LARGE SAMPLE TESTS OF IDENTIFICATION AND SPECIFICATION IN THE LINEAR SIMULTANEOUS EQUATIONS MODEL

by

Leonard Gill

Ph.D.

University of Edinburgh

1983

I declare that I have solely composed this thesis.

....

I wish to thank:

my supervisors, Peter Fisk and Peter Vandome, for their help and support over the years;

my colleagues in Manchester, for their patience and encouragement;

Dr. John Smith, whose graphic text formatter and character set editor made the printing of the thesis possible;

lastly, but by no means least, my wife and family, without whose support I would have given up long ago.

CONTENTS

Chapter 1. Introduction. Page

1.1. Introduction and Overview 1

1.2. Inference in the Linear Simultaneous Equations Model 5

1.3. Some Specific Hypotheses 9

1.4. A More Detailed Outline 23

1.5. Notation and Conventions 34

1.6. Some Useful Results 39

Appendix to Chapter 1 50

Chapter 2. Estimation and Inference for Constraint Parameter Problems

2.1. Introduction 51

2.2. Strong Consistency of Constrained Maximum Likelihood Estimators 55

2.3. First Order Conditions 62

2.4. The Limiting Normal Distribution of the Maximum Likelihood Estimators 66

2.5. The Minimum Chi-Squared Estimator 73

2.6. The Two Step Estimator 78

2.7. A Linearised Minimum Chi-Squared Estimator 85

2.8. The Likelihood Ratio Statistic 88

2.9. The Lagrange Multiplier Statistic 93

2.10. Wald Test Statistics 98

2.11. The C-alpha Statistic 103

2.12. Difference Statistics 107

2.13. An Overview and Evaluation 113

Chapter 3. Estimation and Identification in the Linear Simultaneous Equations Model

3.1. Introduction and Outline 119

3.2. Identification 125

3.3. Estimation in the Linear Simultaneous Equations Model 130

3.4. Asymptotic Properties of the Maximum Likelihood Estimators 142

3.5. Two Step Estimation 153

3.6. Minimum Chi-Squared Estimators 165

3.7. The Limited Information Maximum Likelihood Estimator 170

3.8. Conclusions 190

Chapter 4. Tests of the Conditions for Identification

4.1. Introduction 192

4.2. Tests of Identification: A Survey 197

4.3. Sample Criteria for Identification 203

4.4. Limit Distributions of Sample Rank Criterion Matrices 207

4.5. The Limiting Distribution of the Characteristic Roots of Random Symmetric Matrices 216

4.6. Tests on the Rank of $R_n(\pi_0)$ and $R_n^*(\pi_0)$ 239

4.7. Tests of Consistency 254

4.8. Some Comparisons and Interpretations 261

Appendix to Chapter 4 267

Chapter 5. Tests of Overidentifying Restrictions

5.1. Introduction 270

5.2. Tests of Overidentifying Restrictions in the Simultaneous Equations Model: A Survey 277

5.3. Likelihood Ratio Tests of Overidentifying Restrictions 295

5.4. The Lagrange Multiplier Test Statistic 301

5.5. The C-alpha Test Statistic 305

5.6. Wald Test Statistics 310

5.7. Difference Statistics 318

5.8. Limited Information Tests 323

5.9. Conclusions 333

Chapter 6. Estimation and Inference in Unidentified Models

6.1. Introduction 341

6.2. Large Sample Arguments in an Unidentified Model 353

6.3. Estimation in an Unidentified Model 363

6.4. Inference in an Unidentified Simultaneous Equations Model 373

6.5. Conclusions 388

Chapter 7. Tests of Non-Nested Hypotheses: A General Discussion

7.1. Introduction 393

7.2. Interpreting the Results of a Test of Non-Nested Hypotheses 401

7.3. Test Statistics for Non-Nested Hypotheses:

A Survey 405

7.4. Conclusion 426

Chapter 8. Tests of Restricted Non-Nested Hypotheses

8.1. Introduction 428

8.2. The Basic Distributional Results 430

8.3. Cox-Type Statistics for Restricted Non-Nested Hypotheses 449

8.4. Encompassing Test Statistics 458

8.5. Two Step Estimators and Tests of Non-Nested Hypotheses 466

8.6. Tests of Non-Nested Hypotheses in a Certain Exponential Family of Distributions 479

8.7. Conclusion 495

Chapter 9. Tests of Non-Nested Hypotheses in the Linear Simultaneous Equations Model

9.1. Introduction 501

9.2. Limit Distributions for Tests of Non-Nested Hypotheses in the Linear Simultaneous Equations Model 514

9.3. Cox-Type Statistics 530

9.4. Wald and Score Difference Encompassing Test Statistics 542

9.5. Some Special Cases 561

9.6. Conclusions 584

Chapter 10. Overview and Criticism

10.1. Motivation and Evaluation 589

10.2. Some Critical Observations 598

10.3. Some Suggestions for Further Research 602

References 604

UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS (Regulation 3.9)

Name of Candidate Leonard Gill

Address

Degree Ph.D *Date* 1983

Title of Thesis Large Sample Tests of Identification and Specification in the Linear Simultaneous Equations Model.

This thesis considers tests of identification and specification in the linear simultaneous equations model, using the results of full information maximum likelihood estimation of the parameters, by viewing the simultaneous equations model as a special case of a general constrained maximum likelihood problem, in which the parameter restrictions are expressed in "freedom equation" or "constraint parameter" form.

In analysing this general problem, a new type of Wald test statistic based on minimum chi-squared estimation is developed, having a property of symmetry, in that knowledge of the structural parameter restrictions on the alternative hypothesis model is not required to construct the test statistic.

The identification of the parameters of the linear simultaneous equations model depends on the satisfaction of two conditions, the "rank condition" and the "consistency condition": tests of null hypotheses of the satisfaction of these conditions are constructed using a limit normal distribution for the smallest characteristic roots of certain random symmetric matrices. A one-sided confidence bound procedure is used to construct a test of the rank condition, whilst a standard one-sided large sample test is used for the consistency condition.

Tests of overidentifying restrictions in a linear simultaneous equations model are considered by specialisation from the general results mentioned earlier, but the structure given to the test statistics by the nature of the simultaneous equations model is exploited to find useful ways of calculating the values of the test statistics. In this particular discussion, it is assumed that the parameters of the null and alternative hypothesis models are identified. This assumption is later relaxed, to allow a situation in which the parameters of each model are unidentified. Useful asymptotic results are obtained by the imposition of additional identifying restrictions, which may be regarded from one viewpoint as "true", and from another, "arbitrary". Consideration is then given to various notions of invariance of test statistics for additional restrictions imposed on the alternative hypothesis model: not all of the test statistics discussed satisfy all of the notions of invariance considered.

Finally, tests of non-nested hypotheses in a linear simultaneous equations model are considered. As a general problem, this requires the construction of test statistics for restricted non-nested hypotheses, from which test statistics for the simultaneous equations model can be obtained by specialisation. Initial consideration is given to two types of statistic, "Cox-type", and "Encompassing-type" statistics, and then a family of "score difference encompassing test" statistics is constructed, which are quite easy to calculate. A variety of special cases of the simultaneous equations model are considered, and their impact on the test statistics investigated.

Chapter 1 : Introduction

1.1. Introduction and Overview

The arguments advanced in this thesis are very much in the spirit of recent developments in the use of parametric hypothesis tests in econometrics. This spirit may be described as the detection of misspecification of all kinds in a postulated model (i.e. "misspecification tests"), and the counterpart, the verification of parameter restrictions implied by the (economic) theory underlying the postulated model (i.e. "specification tests").

1.1.1. One may describe misspecifications as being failures in the statistical assumptions underlying the postulated model: for example, failure of independence or homoscedasticity of error terms, functional form misspecification, omission of relevant variables. However, one could also describe the imposition of incorrect parameter restrictions as being a misspecification, so that a specification test may also be used as a misspecification test, at the user's discretion.

Typically, specification and misspecification tests view the null hypothesis to be tested as a more restricted or specialised version of the model assumed to be true under the alternative hypothesis: this type of test situation is one of nested hypotheses, where the differences between the two hypotheses can be expressed solely by differences in the values of certain parameters. One can thus see that as the nature of the postulated model varies, the status of a particular test as a specification or misspecification test may change. For example, serial correlation is frequently regarded as a nuisance in regression models, yet it is the lifeblood of time series models: thus, investigations of the nature of the structure of such serial correlation have different emphases which roughly match the misspecification-specification test distinction.

Broadly speaking, the uses made of the inferential procedures above may be described as either "learning from the data", or "confirming one's theory": quite often, a test statistic designed for a specific purpose is employed by a researcher to detect a more general kind of deviation away from the null hypothesis model. A good example of this is the well-known Durbin-Watson statistic, for testing independence against an AR(1) model for the errors, which is widely used as an indicator of general serial correlation, omitted variables, or misspecification of the dynamics of the equation being estimated.

The well known Likelihood Ratio (Neyman and Pearson [1928]), Wald (Wald [1943]), and Lagrange Multiplier (Rao [1948], Aitchison and Silvey [1958,1959]) test principles for such nested hypotheses have been widely applied in econometrics; these statistics will be briefly described in subsection 1.3.3. . A common view appears to be that the Lagrange Multiplier test principle is well suited to tests of misspecification, whilst the Likelihood Ratio and Wald tests are more suited to be tests of specification: see for example Engle [1982]. The grounds for such an argument seem to be that the Lagrange Multiplier statistic only requires estimation of the model under the null hypothesis, the Wald statistic requires estimation only under the alternative hypothesis, whilst the Likelihood Ratio statistic requires estimation under both hypotheses. Precise specification of the null and alternative hypothesis models is required for each of the three test statistics, but it often turns out that the given Lagrange Multiplier statistic is valid against wider hypotheses than the specified one. This would seem to justify the classification of the three types of test statistic. As an example, the Lagrange Multiplier test of independence against an AR(p) alternative model in a static regression model is also valid against MA(p) alternatives.

Another recent development has been the construction of test statistics for hypotheses that are non-nested: that is, they cannot be regarded as a special case, or restricted version, of the alternative hypothesis; a more precise definition will appear later in the chapter. Such situations can arise in a number of ways: the typical one is where the substantive theories embodied in the competing statistical models are conflicting. Another aspect of this concerns the embedding of such competing models in a more general model, i.e. so that they become, individually, nested special cases: one can point to circumstances where this is not possible, or simply unsatisfactory from the point of view of the substantive theory. More detailed discussion of this point will be given in Chapter 7.

There are technical reasons why the Likelihood Ratio, Wald and Lagrange Multiplier test statistics cannot be used directly for such non-nested hypotheses; instead, one has to modify these statistics in various appropriate ways. It has been argued above that specification tests are concerned with detecting further specialisations within a given model; if this is accepted, tests of hypotheses that are non-nested would seem, almost by definition, to be misspecification tests. This is not a universally accepted view, however; some investigators regard non-nested hypothesis tests as ways of choosing between the null and alternative hypothesis models. This viewpoint is examined more fully in Chapter 7.

1.2. Inference in the Linear Simultaneous Equations Model

1.2.1. In this thesis, a number of closely related topics in estimation and inference in the linear simultaneous equations model are considered. A certain amount of generality is given by allowing the structural parameter restrictions to be across-equation, linear and inhomogeneous, together with the use of maximum likelihood methods to estimate all the free parameters of the model.

This estimation problem is attacked indirectly by embedding the simultaneous equations model in a more general restricted maximum likelihood estimation problem. There are a number of advantages in this approach : one has to use structural parameter restrictions in the general framework, and a number of interesting, if possibly well-known, results follow from this, specifically with regard to the nature of the three test statistics, Likelihood Ratio, Lagrange Multiplier, and Wald; these statistics are denoted LR, LM and W for short. In addition, the literature on non-nested hypothesis tests has so far concentrated on such a general framework, where the parameters are unrestricted. A discussion of Lagrange Multiplier and Wald statistics in the simultaneous equations model is facilitated by this embedding, since they are usually, if not exclusively, associated with maximum likelihood estimation. Finally, it is felt that the "full information maximum likelihood" estimator in the simultaneous equations model is of some interest in itself.

1.2.2. In order to discuss the types of hypotheses in the simultaneous equations model for which test statistics will be derived, it is necessary to define the linear simultaneous equations model: it can be regarded as a linear relationship between the $m \times 1$ random vector y_t and a $k_1 \times 1$ vector of nonstochastic and linearly independent exogenous variables x_t, observed at points $t = 1,\ldots,n$:

$$y_t = \Pi_1' x_t + v_{1t},$$

usually called the "reduced form", where v_{1t} is an unobserved error term with mean vector 0 and covariance matrix Ω_1; Π_1 is a $k_1 \times m$ matrix of parameters generated from the "structural form" of the model:

$$A_1' y_t + B_1' x_t = u_t,$$

where A_1 and B_1 are $m \times m$ and $k_1 \times m$ matrices of unknown parameters, A_1 being nonsingular, u_{1t} having mean vector 0 and covariance matrix

$$\Sigma_1 = A_1' \Omega_1 A_1 .$$

By constructing "observation matrices"

$$Y' = (y_1,\ldots,y_n), \quad X' = (x_1,\ldots,x_n),$$

$$U_1' = (u_{11},\ldots,u_{1n}), \quad V_1' = (v_{11},\ldots,v_{1n}),$$

of respective dimensions $m \times n$, $k_1 \times n$, $m \times n$, and $m \times n$, one can write

$$Y = X\Pi_1 + V_1, \qquad \langle 1.2.2.1\rangle$$

or

$$YA_1 + XB_1 = U_1. \qquad \langle 1.2.2.2\rangle$$

Let

$C_1' = (A_1' \vdots B_1')$,

with

$C_1 = (c_{1.1}, \ldots, c_{1.m})$,

an $(m+k_1)\times m$ matrix, and

$g_1' = (c_{1.1}', \ldots, c_{1.m}')'$.

In subsection 1.6.1, such a "long vector" g_1 is defined as the "vec" of the matrix from which it is constructed: so,

$$g_1 = \text{vec } C_1; \qquad \text{<1.2.2.3>}$$

see also subsection 1.6.1. .

The vector g_1 contains all the elements of A_1 and B_1: i.e. the structural form parameters; these satisfy certain restrictions, here described by the linear inhomogeneous equations

$$g_1 = K\delta + k, \qquad \text{<1.2.2.4>}$$

where K is a known $m(m+k_1)\times q_1$ matrix with full column rank, and k a known $m(m+k_1)\times 1$ vector. The unrestricted structural parameters are contained in the vector δ. There is a basic notational difficulty here: for estimation purposes, it would be quite satisfactory to dispense with the subscripts "1" on A, B, C, Π, etc; however, when one has to estimate "different" models under a null and an alternative hypothesis, it is convenient to use subscripts "$_0$" and "$_1$" for this purpose. So, the above is a description of the alternative hypothesis model.

This particular formulation is quite a natural generalisation of the traditional "within equation exclusion restrictions and unit normalisation" case discussed in the textbooks: in this special case, the vector k would have zero elements except for units in the $[(i-1)(m+k_1)+i]$th positions, $i = 1,\ldots,m$, whilst δ would consist of those parameters remaining in the structural equations after the restrictions have been imposed.

It should be pointed out that the vector of "exogenous" variables x_t is not allowed to contain any lagged dependent variables, mainly for simplicity: in general, the large sample results will carry over to dynamic models, provided that the error vectors u_t are independent and that the model is covariance stationary for y_t, but not stationary in the mean.

1.3. Some Specific Hypotheses

A more detailed outline of the contents of the remaining chapters will be given below; for the moment, a sketch of the nature of the tests of hypotheses discussed in this thesis will now be given.

1.3.1. The parameter vector δ of equation <1.2.2.4> above is said to be identified if and only if there is a unique solution to the equation

$(I_m \otimes (\Pi_1 : I_{k_1}))K\delta = (I_m \otimes (\Pi_1 : I_{k_1}))k$:

that is, if and only if a solution exists and

$(I_m \otimes (\Pi_1 : I_{k_1}))K$

has full column rank $m(m+k_1)$. This latter condition is called the "rank condition" for identification.

A lack of identification, given a set of a priori structural restrictions, is a serious problem, since an investigator has no means of knowing which out of a number of possible structures generated the specified reduced form parameter values Π_1, Ω_1, and hence, no fixed structural parameter values on which to base theoretical or economic inferences from the structural specification of the model. It is argued in Chapter 4 that apparently well determined structural parameter estimates may suffer (statistically) from a near failure of identification: if this occurs, the natural conclusion to draw is that the sample data is "nearly" compatible with different structural parameter values. This situation is analogous to the problem of near-multicollinearity in linear regression models.

It thus seems reasonable to suppose that a sample criterion, on which a test of identification could be based, would be a useful empirical tool: the matrix

$$(I_m \otimes (\Pi_1 : I_{k_1}))K$$

depends on Π_1 which can be estimated, so that an estimate of the rank of this matrix can be obtained by estimating the characteristic roots of matrices like

$$K'(I_m \otimes \begin{bmatrix} \Pi_1' \\ I_{k_1} \end{bmatrix} X'X(\Pi_1 : I_{k_1}))K$$

or

$$K'(\Sigma^{-1} \otimes \begin{bmatrix} \Pi_1' \\ I_{k_1} \end{bmatrix} X'X(\Pi_1 : I_{k_1}))K.$$

The difficulty is then in finding a limiting distribution for these characteristic root estimates on which to base a test statistic.

It will be argued in Chapter 4 that the appropriate null hypothesis is that the model is identified, with the alternative being that of a lack of identification, and that anything other than clearcut evidence in favour of the null hypothesis will lead to its rejection: the test proposed there is perhaps a more informal type of significance test than the types of test based on nesting arguments so far discussed. This suggests that the test be viewed as a rather weak test of specification: one desires to check that the structural model is identified before proceeding to investigate other aspects of its structure; or, more crudely, that the data will tolerate the structure the investigator wishes to impose. It should be pointed out that tests of identification of the null hypothesis model are considered in Chapter 4: attention is now turned to defining this model.

1.3.2. It has already been indicated that results on estimation and inference in the simultaneous equations model will be obtained by embedding it in a general maximum likelihood problem: thus, it will be necessary to sketch this framework, and introduce some necessary notation.

It is assumed that the observable independent random vectors $y_1, \ldots, y_n$, arranged in a long vector

$y' = (y_1', \ldots, y_n')$

have the scaled log-likelihood function

$n^{-1} l_n(y; \theta)$,

dependent on the $s_0 \times 1$ parameter vector θ: under the null hypothesis,

$\theta = \theta(\alpha)$,

α an $r_0 \times 1$ vector belonging to the parameter space $\mathcal{A}$, whilst under the alternative hypothesis,

$\theta = \phi(\beta)$,

β an $r_1 \times 1$ vector belonging to the parameter space $\mathcal{B}$. The restrictions

$\theta = \theta(\alpha)$

are regarded as arising from

$$\theta = \phi(\beta)$$

by further restriction: i.e. the null hypothesis could be written as

$$\theta = \phi(\beta), \quad \beta = \lambda(\alpha),$$

or

$$\theta = \phi[\lambda(\alpha)].$$

This is called the "constraint parameter" or "freedom equation" formulation: α and β are regarded as the free parameters of each hypothesis, and are assumed to be (at least, locally) identified.

The maximum likelihood estimators under each hypothesis are then obtained by maximising

$$n^{-1}l_n(y;\theta)$$

subject to

$$\theta = \theta(\alpha)$$

for the null hypothesis model, and maximising

$$n^{-1}l_n(y;\phi)$$

subject to

$$\phi = \phi(\beta)$$

for the alternative hypothesis model: the estimates are denoted $\tilde{\theta}$, $\tilde{\alpha}$ and $\tilde{\phi}$, $\tilde{\beta}$ respectively. The notational switch in this is more apparent than real: $\tilde{\phi}$ is an estimator of the true value of θ when the null hypothesis is true, and the true value of ϕ when the alternative is true. This type of notational problem has been mentioned before, and arises simply because of the desire to consider both issues of estimation and inference in an integrated way. The true value of a parameter under a hypothesis will be given a superscript "^{0}": so, θ^0 and α^0 are the true values under the null hypothesis, and similarly, β^0, ϕ^0 are the true values of β and ϕ under the null hypothesis.

A test of these hypotheses is clearly a test of the restrictions embodied in

$\beta = \lambda(\alpha)$.

There exists a function $f(.)$ such that

$f(\beta) = 0 \quad$ if and only if $\beta = \lambda(\alpha)$,

(at least locally, by the implicit function theorem), so that the truth of the additional restrictions on β,

$f(\beta) = 0$,

is being tested.

When θ and β have the same dimension ($r_1 = s_0$), and to each θ there corresponds a unique β, one can say that the restrictions

$\theta = \phi(\beta)$

"just-identify" β. In econometric terminology, the restrictions

$\beta = \lambda(\alpha)$

serve to "overidentify" β, in that more restrictions have been imposed than are necessary to identify β. When $r_1 < s_0$, the alternative hypothesis model can then be described as "overidentified", in that it could be obtained by imposing additional restrictions on a just-identified model.

Tests of the hypotheses

$$H_0: \quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha)$$

$$H_1: \quad \theta = \phi(\beta)$$

using the Likelihood Ratio, Lagrange Multiplier, Wald and other statistics are constructed in Chapter 2; here, a brief sketch of the nature of the Likelihood Ratio and Lagrange Multiplier statistics is given. The Likelihood Ratio statistic is given by

$$LR = -2[l_n(y;\tilde{\phi}) - l_n(y;\tilde{\theta})]$$

and will have, under appropriate assumptions, a limit χ^2 distribution with degrees of freedom equal to the difference in the dimensions of β and α,

$$r_1 - r_0,$$

under the null hypothesis. The Lagrange Multiplier statistic is based on the estimated Lagrange multiplier, $\tilde{\tau}$ of the null hypothesis model, or, equivalently, the estimated score vector,

$$D_\theta l_n(y;\tilde{\theta})$$

and has the same limit distribution on the null hypothesis as the Likelihood Ratio statistic. Conventionally, Wald test statistics focus on the relationship $\beta = \lambda(\alpha)$ expressed in the constraint equation form,

$$f(\beta) = 0,$$

and examine whether $f(\tilde{\beta}) = 0$; it is not easy to see how to construct a Wald statistic in the constraint parameter framework without resorting to a conversion to a constraint equation framework, but a suitable method for this is suggested in Chapter 2, and briefly described in subsection

1.4.1. following.

1.3.3. To embed the simultaneous equations model in this framework, it is necessary to regard the simultaneous equations model described in equations <1.2.2.1>–<1.2.2.4> as that holding under the alternative hypothesis. Under the null hypothesis, the free parameter δ is regarded as being further restricted:

$\delta = L\gamma + r$,

where L is a known $q_1 \times q_0$ $(q_0 < q_1)$ matrix of full column rank, γ a $q_0 \times 1$ vector of free parameters, and r a known $q_1 \times 1$ vector. However, it is convenient to write out explicitly the simultaneous equations model that is supposed to rule under the null hypothesis, the structural form being

$A_0' y_t + B_0' x_t = u_{0t}, \quad t = 1,\dots,n,$

A_0 an $m \times m$ nonsingular matrix, and B_0 a $k_1 \times m$ matrix. Let

$U_0' = (u_{01},\dots,u_{0n})$:

then, as before, one can write

$$YA_0 + XB_0 = U_0. \qquad \langle 1.3.3.1\rangle$$

The reduced form is

$y_t = \Pi_0' x_t + v_{0t}, \quad t = 1,\dots,n,$

or,

$$Y = X\Pi_0 + V_0, \qquad \langle 1.3.3.2\rangle$$

where

$V_0' = (v_{01},\dots,v_{0n})$.

The covariance matrices of the zero mean random vectors u_{0t} and v_{0t} are Σ_0 and Ω_0, which are connected by the relationship

$\Sigma_0 = A_0'\Omega_0 A_0$.

Let

$$C_0' = (A_0' : B_0'),$$

with

$$C_0 = (c_{0.1}, \ldots, c_{0.m});$$

the arrangement of the columns of this matrix in a long vector, defined as

$$g_0 = \text{vec } C_0 = \text{vec}(c_{0.1}, \ldots, c_{0.m}) \qquad \langle 1.3.3.3\rangle$$

enables the structural parameter restrictions to be written as

$$g_0 = H\gamma + h, \qquad \langle 1.3.3.4\rangle$$

where H is a known $m(m+k_1)\times q_0$ matrix of full column rank, h a known $m(m+k_1)\times 1$ vector, and γ the $q_0\times 1$ vector of free structural parameters.

The parameter vectors of the general problem, θ and ϕ, are then taken to consist of reduced form parameters: specifically, vec Π_0 and vec Π_1, and the distinct elements of the covariance matrices Ω_0 and Ω_1, arranged in lexicographic order in the vectors $v(\Omega_0)$ and $v(\Omega_1)$; more details on such vectors are given in subsection 1.6.3.. Thus,

$$\theta = \begin{bmatrix} v(\Omega_0) \\ \text{vec } \Pi_0 \end{bmatrix}; \quad \phi = \begin{bmatrix} v(\Omega_1) \\ \text{vec } \Pi_1 \end{bmatrix};$$

to reduce problems of dimensionality, it is convenient to regard the vectors of distinct elements of the two covariance matrices, $v(\Omega_0)$ and $v(\Omega_1)$, as free parameters, so that

$$\alpha = \begin{bmatrix} v(\Omega_0) \\ \gamma \end{bmatrix}, \quad \beta = \begin{bmatrix} v(\Omega_1) \\ \delta \end{bmatrix}.$$

The functional relationship between θ and α, say, is obvious in the case of $v(\Omega_0)$: for vec Π_0 and γ, a composite function relationship arises from writing

$$\text{vec } \Pi_0 = \text{vec } (-B_0A_0^{-1}) = f_0(g_0)$$

and using

$$g_0 = H\gamma + h;$$

here $f_0(.)$ merely serves to express the functional relationship between Π_0 and the elements of C_0.

Given the embedding of the null and alternative hypothesis simultaneous equations models into the general framework, tests of overidentifying restrictions in the simultaneous equations model are then obtained as a special case of tests of the hypotheses

H_0: $\theta = \theta(\alpha)$

H_1: $\theta = \phi(\beta)$,

or equivalently,

H_0: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$

H_1: $\theta = \phi(\beta)$.

It will be shown that this specialisation amounts to a test of

$$\delta = L\gamma + r,$$

which parallels $\beta = \lambda(\alpha)$ in the general case, except that the covariance parameters $v(\Omega_0)$, $v(\Omega_1)$ play no direct role in the test statistics.

Not all of the test statistics discussed focus on this relationship: for example, the Lagrange Multiplier and

Likelihood Ratio statistics for such tests of overidentifying restrictions only involve the reduced form parameter estimates under H_0 and H_1; when β has the same dimension as θ, θ is unrestricted and the maximum likelihood estimator of θ under H_1 can correspond to any "just-identified" version of the structural form (1.2.2.2) on the alternative hypothesis. This naturally leads to interpretations of these two tests as misspecification tests. Ordinarily, a Wald test statistic would focus on the relationship

$$\beta = \lambda(\alpha),$$

and would be thus interpreted as a specification test, since knowledge of the function $\theta = \phi(\beta)$ would be required, but a Wald-type test statistic will be constructed in Chapter 2 that does not depend on the particular function $\phi(\beta)$, so that it can be interpreted as a misspecification test statistic. When the specification

$$\theta = \phi(\beta)$$

is "overidentifying", then all of the tests considered are tests of specification.

<u>1.3.4</u> It is not unreasonable to suppose, in a multi-equation model, that the process of "specification search" may reveal more than one structural specification, each of which is acceptable with respect to the kind of tests discussed above and to other tests of misspecification, but are not nested with respect to each other in the sense discussed earlier. An investigator may then subject his theoretically "preferred" model (if such exists!) to test against one of the

alternative models he is prepared to entertain by means of a suitable test statistic for non-nested hypotheses: that is, even if there are several such alternatives, only pairwise comparisons of the preferred and alternative models are to be made. It might be argued that in such circumstances, it would be better to attempt a "synthesis" of the competing models within a more general model, but as noted earlier, this may not always be desirable or feasible.

Such non-nested test statistics for simultaneous equation models will be generated by specialisation of test statistics obtained for a more general problem: it will be convenient for this introduction to subsume the data density for the independent random vectors $y_1,\ldots,y_n$ under each competing hypothesis into the corresponding log-likelihood functions. As far as is feasible, the notation used for the general nested test problem is used in the non-nested test case: in general, however, the log-likelihood under the null hypothesis,

$l_n(y;\theta)$

differs from that under the alternative, which is supposed to be

$m_n(y;\phi)$,

with no necessary relation between θ and ϕ. So, more formally,

H_0: the log-likelihood is $l_n(y;\theta)$, and $\theta = \theta(\alpha)$;

H_1: the log-likelihood is $m_n(y;\phi)$, and $\phi = \phi(\beta)$.

For the simultaneous equations model, the model supposed to rule under the alternative hypothesis H_1 is that given in equations <1.2.2.1>-<1.2.2.4>, whilst on the null hypothesis, there may be a completely different regressor set W, as well as different structural form restrictions: the reduced form may be written as

$$y_t = \Pi_0' w_t + v_{0t}, \qquad t = 1,\ldots,n,$$

where w_t is a $k_0 \times 1$ vector of non-stochastic and linearly independent exogenous variables; in observation matrix form, the reduced form is

$$Y = W\Pi_0 + V_0,$$

with

$$W' = (w_1,\ldots,w_n).$$

The structural form is

$$A_0' y_t + B_0' w_t = u_{0t}, \qquad t = 1,\ldots,n,$$

or,

$$YA_0 + WB_0 = U_0; \qquad \text{<1.3.4.1>}$$

defining

$$C_0' = (A_0' : B_0'),$$

$$g_0 = \text{vec } C_0, \qquad \text{<1.3.4.2>}$$

the structural restrictions are

$$g_0 = H\gamma + h, \qquad \text{<1.3.4.3>}$$

just as in the nested case. However, here, it is not supposed that there exists a matrix L and a vector r to connect δ and γ via

$$\delta = L\gamma + r.$$

A number of interesting special cases of these two competing models are considered: where there is the same exogenous variable set X under both the null and alternative hypotheses, but different sets of structural restrictions. Provided that these restrictions are sufficiently different to prevent degeneracy into a test of additional overidentifying restrictions, one can argue that at least some of the restrictions under the null hypothesis serving to identify that model are being subjected to test: the details of this argument are somewhat complex, and further discussion is postponed until Chapter 9.

There is quite a large variety of possible test statistics for non-nested hypotheses, some being modifications of the Likelihood Ratio statistic, and others closely related to the usual Wald and Lagrange Multiplier statistics: a detailed examination of all these test statistics is given in Chapter 8. It is interesting that the Wald and Lagrange Multiplier statistics have the property of collapsing to their usual form when in fact the null hypothesis can be nested in the alternative hypothesis.

<u>1.3.5</u> All of the test statistics derived in the pages that follow have only large sample validity: an obvious criticism of such statistics is that it is not known in general how good are the limiting distributions as approximations to the unknown finite sample distributions. It seems natural to say that exact finite sample distribution results are more desirable, except for one thing: such results usually depend on a normality assumption for the underlying observations,

and one does not usually know how well this assumption is satisfied, unless a test for normality is made. Usually, a normality assumption is made in this thesis, although most of the large sample distribution results do hold without such an assumption; the test statistics for non-nested hypotheses are possibly an exception to this conclusion.

1.4. A More Detailed Outline.

The rest of this chapter contains a reasonably detailed outline of the contents of each chapter of this thesis, with an indication of which material is thought to be new or novel, followed by a discussion of some of the notational principles and conventions adhered to. The final section of this chapter consists of some useful mathematical and statistical results employed in later chapters.

1.4.1. In Chapter 2, the maximum likelihood estimators of the parameters θ and β are found for the problem where $y_1, \ldots, y_n$ are observable random vectors with respective densities $f_t(y_t;\theta)$, $t = 1,\ldots,n$, but in addition, it is known that $\theta = \phi(\beta)$.

The strong consistency, under the null hypothesis, of these maximum likelihood estimators, $\tilde{\phi}$ and $\tilde{\beta}$ for the true values θ^0 and β^0 is established under the assumption of compactness of the parameter space of β and some uniform convergence assumptions. It is also shown that the estimated Lagrange Multiplier, $\tilde{\mu}$, converges almost surely to 0. A joint limiting normal distribution for

$$n^{1/2}(\tilde{\phi} - \theta^0), \quad n^{1/2}(\tilde{\beta} - \beta^0), \quad \text{and} \quad n^{1/2}\tilde{\mu}$$

is also obtained. Two alternative estimation methods, "minimum chi-squared" and "two-step", are also examined under the same circumstances.

Given these results, test statistics are constructed for the case in which the model above describes the alternative hypothesis, and where under the null hypothesis β is further restricted by

$$\beta = \lambda(\alpha),$$

so that on the null hypothesis,

$$\theta = \phi(\lambda(\alpha)) = \theta(\alpha).$$

Separate consideration is given to the case where the dimensions of θ and β are the same, and $\phi(\beta)$ has a unique inverse function, so that $\tilde{\phi}$ is actually the unrestricted maximum likelihood estimator. This leads into a discussion of "symmetric" and "asymmetric" test procedures: a symmetric test procedure is one where the investigator does not need to specify the function

$$\phi(\beta)$$

explicitly, and is therefore unable to state what the "structure" of the model is when the null hypothesis fails. An asymmetric test requires the specification of $\phi(\beta)$ as part of the alternative hypothesis.

The test statistics considered are the Likelihood Ratio (LR), the Lagrange Multiplier (LM), the Wald (W) and the C-alpha, usually denoted $C(\alpha)$, (CA); this statistic is due to Neyman [1959], and is defined more formally in Chapter 2. Of these statistics, it is usually thought that in the case where θ is unrestricted on the alternative hypothesis, the Likelihood Ratio, Lagrange Multiplier, and C-alpha tests are symmetric, whilst the Wald test is asymmetric in the sense used above. However, it is shown how to construct a new symmetric Wald test statistic, using the minimum chi-squared estimation principle. When the alternative hypothesis is restricted, all of the test statistics are asymmetric.

<u>1.4.2.</u> In Chapter 3, prior to considering estimation of the parameters of the simultaneous equations model described by equations <1.2.2.1>-<1.2.2.4>, the conditions for the identification of the structural parameters are considered. The parameters of this model are then estimated by maximum likelihood, assuming normality of the structural and reduced form error terms: from the first-order conditions, an "almost" explicit form for the estimator of δ is obtained, similar to that obtained by Hendry [1976]; the estimator is called the "full information maximum likelihood" (FIML) estimator. The formal limiting distribution of all the parameter estimates is then obtained.

Given all this information, the nature of a "two-step" estimator of the parameter vector δ is considered: this leads on naturally to the question of whether such estimators can be obtained directly by a regression, rather than indirectly via an "update" term. Some conditions for such an estimator to be asymptotically efficient given by Hendry [1976] are investigated.

Hendry [1976] and others have shown that the limited information maximum likelihood (LIML) estimator of the parameters of a single structural equation may be obtained from the FIML estimator of a simultaneous equations model with a single over-identified structural equation, and the remaining $m-1$ equations being in reduced form, with no across equation restrictions. This argument is then used to obtain the LIML estimator for a single over-identified structural equation, with general within-equation, linear inhomogeneous restrictions of the form

$$g_{1.1} = K_{11}\delta_{.1} + k_{.1},$$

where

$$g_1' = (g_{1.1}', \ldots, g_{1.m}'),$$
$$\delta' = (\delta_{.1}', \ldots, \delta_{.m}')$$
$$k' = (k_{.1}', \ldots, k_{.m}'),$$

and $g_{1.1}$, $\delta_{.1}$, K_{11} and $k_{.1}$ refer to the first structural equation, which is assumed to be the over-identified one. It is believed that this generalisation of the usual framework for the LIML estimator of only within-equation exclusion restrictions and a unit normalisation rule has not yet appeared in the literature. The results can be specialised to produce the standard LIML estimator for this case.

1.4.3. A new approach to tests of the identification of the parameters of a simultaneous equations model is considered in Chapter 4; it is based on the use of the rank of estimates of matrices of the form

$$H'(I_m \otimes \Omega_0'X')N(I_m \otimes X\Omega_0)H,$$

where N is positive definite, to estimate the rank of the matrix

$(I_m \otimes X\Pi_0)H$ or $(I_m \otimes \Pi_0)H$

which appears in the rank condition for identification of the structural parameter vector γ of the model defined by equations <1.3.3.1>-<1.3.3.4>.

The choices of N considered are $\tilde{\Sigma}_0^{-1} \otimes I_n$, where $\tilde{\Sigma}_0$ is the maximum likelihood estimator of Σ_0, together with the maximum likelihood estimator of Π_0, and $I_m \otimes I_n$, where the estimator of Π_0 is then the ordinary least squares estimator. These choices correspond to post- and pre-estimation criteria for identification respectively; the limiting distribution of $n^{-\frac{1}{2}}$ times these criterion matrices is found under a null hypothesis of identification, and an argument given by Anderson [1963] is adapted to find a corresponding limiting normal distribution for their characteristic roots; this derivation is somewhat complex. Specialisation of the limiting distributions is made to the case of within-equation exclusion restrictions and unit normalisation rules.

A one-sided confidence interval approach is used to provide a method of dealing with the essential non-negativity of the characteristic roots of the population criterion matrix $H'(I_m \otimes \Pi_0'X')N(I_m \otimes X\Pi_0)H$.

In deriving the rank test as a test of identification, it is assumed that there exists a solution to the equation

$(I_m \otimes X\Pi_0)H\gamma = -(I_m \otimes X\Pi_0)h$;

this assumption is also tested, given the satisfaction of the rank condition, using a test on the smallest roots of a criterion matrix similar to that used in the rank test: this is called a "consistency test".

For the case of a model consisting of a single overidentified equation, with the remaining equations in reduced form, analogies are drawn between the rank and consistency tests described above, and the well-known LIML-based single and double root tests: it is concluded that given that the rank condition holds, the consistency test is essentially a test of the overidentifying restrictions embodied in the first equation, but not of the classical kind discussed in the following chapter.

1.4.4 In Chapter 5, the general inference results of Chapter 2 are specialised to the case of the simultaneous equations model, using the results of Chapter 3. The construction of tests of the structural parameter restrictions is preceded by a critical survey of the literature, which points up a number of confusions to be found there. Particular attention is paid to the construction of test statistics for the various hypotheses considered that are comparatively easy to calculate: the main interest in this chapter lies in revealing the similarities and contrasts between the various test statistics.

Finally, test statistics based on LIML estimation (regarded as a special case of the FIML estimator) for the situation described at the end of subsection 1.4.2 above are obtained; this specialisation is shown to coincide with that obtained from the Anderson and Rubin [1949] formulation of the problem. The Lagrange Multiplier statistic in either formulation turns out to be n times a certain "smallest" characteristic root; a C-alpha test statistic is also examined. It is believed that some of the test statistics suggested in this chapter have not yet appeared in the literature explicitly.

1.4.5. One interesting novelty of the thesis is an examination of the possibility of inference on the parameters of an unidentified simultaneous equations model, which is described in Chapter 6. A basic result on the almost sure convergence of any solution to the maximum problem

$\max_{\phi} \quad n^{-1} l_n(y;\phi)$

subject to

$\phi = \phi(\beta)$

is proved, for the case where β is not necessarily identified; the proof is based on elementary arguments, in contrast to those given by Redner [1981], who considered a related problem. The results are then used to justify the use of generalised inverses in obtaining a limiting joint normal distribution for

$n^{1/2}(\tilde{\phi} - \phi^0)$, $n^{1/2}(\tilde{\beta} - \beta^*)$ and $n^{1/2}\tilde{\mu}$,

where ϕ^0 is the true value of ϕ, β^* a member of the set of solutions to

$\phi^0 = \phi(\beta)$,

and $\tilde{\mu}$ a Lagrange multiplier arising from the constrained maximum problem.

As a preparation for the application of these results to the simultaneous equations model, the role of "estimable", or equivalently, "identifiable" linear functionals of the structural parameter vector δ is investigated, together with a discussion of the mutual relationship of structures that are described as "just-unidentified" and "over-unidentified". That is, unidentified structures that would none the less yield unrestricted and restricted reduced form parameters respectively, in the sense used by Rothenberg [1973, p.37].

Tests of all the overidentifying restrictions in an unidentified model or structure are then formally developed, and close consideration is given to the question of whether each suggested test statistic has the same value no matter which solution of the first-order conditions of the FIML estimator is taken. Clearly, without such invariance, different inferences might be obtained from different solutions, which is undesirable.

Whilst the general viewpoint taken in proposing tests of restrictions in such circumstances is to see whether they are possible in principle, further consideration is given to their practicality, particularly with respect to linear estimators like two- or three-stage least squares.

1.4.6. Chapter 7 consists of a critical survey of the literature on the nature of non-nested hypotheses from a general point of view, and of the test statistics that have been proposed. A sketch is given of the nature of the limiting distributions of these test statistics: the technical details are relegated to Chapter 8, since they are a by-product of the derivation of test statistics for restricted non-nested hypotheses.

1.4.7. The analysis in Chapter 8 examines the large sample arguments required to construct "Cox-type" and "Encompassing-type" test statistics for the two competing hypotheses

H_0: the log-likelihood is $l_n(y;\theta)$ and $\theta = \theta(\alpha)$;

H_1: the log-likelihood is $m_n(y;\phi)$ and $\phi = \phi(\beta)$

for general data densities, and for a sub-family of the exponential family of distributions. The reason for this latter specialisation is again two-fold: considerable simplifications in the form of the various test statistics occur in this specific family, whilst linear normal regression models can be embedded very easily into this family. This approach avoids some (but not all) of the complex algebra which seems inherent in the construction of non-nested test statistics. Some consideration is given to tests based on a "two-step" estimation principle under both the null and alternative hypotheses, as a method of avoiding some of the difficult calculations required to construct certain of the test statistics. It is thought that the development of test statistics for restricted non-nested hypotheses has not yet appeared in the literature.

1.4.8. The application of the analysis of Chapter 8 to the simultaneous equations model is undertaken in Chapter 9: this requires the construction of the limiting distributions, involving some lengthy calculations. Given this, considerable attention is given to methods by which the test statistics can most easily be calculated: in general, these consist of regressions of the structural form or reduced form residual vectors from the estimation of the null hypothesis model on certain regressors, derived from both models. An interesting difficulty arises for certain "Encompassing-type" test statistics in that these regressions will usually be exactly multicollinear, and the number of linearly independent columns of the regressor matrix in these regressions is closely related to the usually unknown degrees of freedom of the limit χ^2 distribution of the test statistics. A method for resolving this problem is given.

A number of special cases are considered, for example, allowing the regressor sets under the null and alternative hypotheses to be the same, allowing one or both of the hypotheses to describe a "just-identified" model, and allowing the null hypothesis to be nested in the alternative in the manner discussed in Chapter 5.

The final issue discussed in Chapter 9 is the possibility that identifying restrictions on a particular simultaneous equations model may be tested via non-nested test statistics: that is, the competing models have the same exogenous variable sets, but different sets of a priori restrictions. Provided that certain constraints on the number of over-identifying and common restrictions between the two models are met, one can interpret the results of a non-nested hypothesis test as being a test of at least some of the identifying restrictions imposed by the null hypothesis model.

1.4.9. In the final chapter, Chapter 10, some conclusions and suggestions for further research are given.

1.5. Notation and Conventions

1.5.1. The general notational principle adopted in this thesis is that each object or concept should have a unique symbol associated with it: to keep to this principle is a very difficult task, even with the use of several different type faces - **bold**, symbol, **Outline**, *ITALIC* and byte. **Bold** symbols, both upper and lower case, will usually refer to quantities from a general model, whilst byte characters will usually refer to quantities in simultaneous equations models. The **Outline** typeface will be used in upper case for "arbitrary" matrices whose meaning will be defined when they are used, whilst lower case **outline** will be split into "**a-g**" for "arbitrary" vectors, "**h-t**" for arbitrary integers, and "**u-z**" for arbitrary variables. This convention will be breached in that **k**,**m**, **q**,**r** and **s**, with or without subscripts, will have a fixed meaning throughout the thesis. The type faces are displayed in full in the appendix to this Chapter.

Some fixed symbols are:-

d : differential

e_i : the ith coordinate vector, dimension to be defined according to context;

n : the sample size;

s : $= - \frac{1}{2}\log 2\pi$;

y : a "generic" random vector;

D : derivative operator;

$I_{\mathbf{p}}$: unit matrix of dimension **p**;

0_{pn} : a $\mathbf{p} \times \mathbf{n}$ zero matrix;

N : the normal distribution – N(a,B);

H : a hypothesis, usually subscripted zero or one.

<u>1.5.2.</u> If $\mathbf{f}(\mathbf{x})$ is a $\mathbf{p} \times 1$ vector function of the $\mathbf{n}$-vector $\mathbf{x}$, then the matrix of partial derivatives is denoted $D_x\mathbf{f}$; the value of this matrix at a point $\mathbf{x}_0$ is

$D_x\mathbf{f}(\mathbf{x}_0)$.

When $\mathbf{p} = 1$, the derivative is written

$(D_x\mathbf{f})'$ or $D_x\mathbf{f}'(\mathbf{x}_0)$,

to emphasise that it is a row vector. The corresponding differential of the function $\mathbf{f}$ is written

$d\mathbf{f} = (D_x\mathbf{f})d\mathbf{x}$

or

$d\mathbf{f} = (D_x\mathbf{f}')d\mathbf{x}$,

according as $\mathbf{p} > 1$ or $\mathbf{p} = 1$.

The $\mathbf{n} \times \mathbf{n}$ second derivative matrix function of the scalar function $\mathbf{g}(\mathbf{x})$ of the $\mathbf{n}$-vector $\mathbf{x}$ will be denoted

$D_x^2\mathbf{g}$ or $D_x^2\mathbf{g}(\mathbf{x})$;

at the point $\mathbf{x}_0$, its value will be denoted

$D_x^2\mathbf{g}(\mathbf{x})$.

The second-order Taylor series expansion of such a scalar function $\mathbf{g}(\mathbf{x})$ around a point $\mathbf{x}_0$ is

$\mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{x}_0) + D_x\mathbf{g}'(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)'D_x^2\mathbf{g}(\bar{\mathbf{x}})(\mathbf{x}-\mathbf{x}_0)$,

where $\bar{\mathbf{x}}$ lies on the line segment having endpoints $\mathbf{x}$, $\mathbf{x}_0$: this will be written as

$\bar{x} \in (x, x_0)$.

The corresponding first-order Taylor series expansion is

$$\mathbf{g}(x) = \mathbf{g}(x_0) + D_x\mathbf{g}'(x^*)(x - x_0),$$

where

$$x^* \in (x, x_0).$$

1.5.3. Quite often, a function of a random vector y and a parameter κ has an expected value function defined by a statistical model for y: in this case, the expected value function of the random vector $\mathbf{f}(y;\kappa)$ will be written as

$$\mathbf{E}_\kappa \mathbf{f}(y;\kappa),$$

or, in a shorthand,

$$[\![\mathbf{E}_\kappa \mathbf{f}]\!];$$

this is a function of κ. A different function of κ is

$$\mathbf{E}_{\kappa^0} \mathbf{f}(y;\kappa),$$

where the expected value operation now uses the specific parameter value κ^0. However, the values of the two functions at κ^0 are identical, namely,

$$\mathbf{E}_{\kappa^0} \mathbf{f}(y;\kappa^0),$$

or, more shortly,

$$[\![\mathbf{E}_\kappa \mathbf{f}]\!]_{\kappa^0}.$$

Expectation and differentiation operators will be combined in a similar way:

$$\mathbf{E}_{\kappa^0}[D_\kappa \mathbf{f}(y;\kappa^0)] = [\![\mathbf{E}_\kappa D_\kappa \mathbf{f}]\!]_{\kappa^0},$$

but note that the functions

$$\mathbf{E}_\kappa D_\kappa \mathbf{f}$$

and

$D_\kappa E_\kappa f$

are in general different.

A similar notation may be used for covariance matrices: the covariance matrix of a random vector y whose distribution depends on a parameter vector κ, as a function of κ, is denoted

$$\mathrm{var}_\kappa(y) = E_\kappa(y - E_\kappa y)(y - E_\kappa y)',$$

but at the fixed parameter point κ^0, it is denoted

$$\begin{aligned}\mathrm{var}_{\kappa^0}(y) &= E_{\kappa^0}(y - E_{\kappa^0}y)(y - E_{\kappa^0}y)' \\ &= [\![E_\kappa(y - E_\kappa y)(y - E_\kappa y)']\!]_{\kappa^0} \\ &= [\![\mathrm{var}_\kappa(y)]\!]_{\kappa^0}.\end{aligned}$$

Similarly, if v and y are jointly distributed random vectors, dependent on a common parameter vector κ, their covariance (matrix) is denoted

$$\mathrm{cov}_\kappa(v,y) = E_\kappa(v - E_\kappa v)(y - E_\kappa y)'$$

whose value at the point κ^0 is denoted

$$\mathrm{cov}_{\kappa^0}(v,y) = [\![\mathrm{cov}_\kappa(v,y)]\!]_{\kappa^0}.$$

1.5.4. Let $\hat{\kappa}$ be an estimator of the parameter vector κ, whose true value is κ^0, and suppose that

$$n^{1/2}(\hat{\kappa}-\kappa^0)$$

has a limit normal distribution with mean vector zero, and some covariance matrix. This limiting covariance matrix will be denoted

$$\Psi(\hat{\kappa};\kappa^0)$$

Ψ denoting "covariance matrix of the limiting normal distribution", the first argument, $\hat{\kappa}$, indicating the estimator concerned, and κ^0 the value of κ at which this matrix function is evaluated. This limiting distribution statement will often be written as

$$n^{1/2}(\hat{\kappa} - \kappa^0) \xrightarrow{d} \mathbf{w},$$

where

$$\mathbf{w} \sim N(0,\ \Psi(\hat{\kappa};\kappa^0)),$$

or, more commonly,

$$n^{1/2}(\hat{\kappa} - \kappa^0) \overset{a}{\sim} N(0,\ \Psi(\hat{\kappa};\kappa^0)).$$

Usually, $\Psi(\hat{\kappa};\kappa^0)$ is the limit of a sequence of matrices, say

$$\{\Psi_n(\hat{\kappa};\kappa^0)\},$$

and an estimator of $\Psi(\hat{\kappa};\kappa^0)$ can often be obtained by replacing κ^0 in $\Psi_n(\hat{\kappa};\kappa^0)$ by the vector $\hat{\kappa}$. So, an estimator of $\Psi(\hat{\kappa};\kappa^0)$ may be denoted

$$\Psi_n(\hat{\kappa}).$$

Sometimes it is necessary to use another estimator of κ^0, say, κ^*: in this case, the estimate of $\Psi(\hat{\kappa};\kappa^0)$ will be denoted

$$\Psi_n(\hat{\kappa};\kappa^*),$$

so that

$$\Psi_n(\hat{\kappa}) = \Psi_n(\hat{\kappa};\hat{\kappa}).$$

1.6. Some Useful Results

1.6.1. If $\mathbf{A}$ is a $\mathbf{p}\times\mathbf{n}$ matrix, the range space or column space of $\mathbf{A}$ is the vector subspace

$$\mathbf{C}(\mathbf{A}) = \{\mathbf{c} \mid \mathbf{A}\mathbf{g} = \mathbf{c}\},$$

whilst the null space or column kernel is

$$\mathbf{N}(\mathbf{A}) = \{\mathbf{g} \mid \mathbf{A}\mathbf{g} = 0\};$$

in each case, $\mathbf{g}$ is $\mathbf{n} \times 1$. It is known that

$$\text{rank } \mathbf{A} = \dim \mathbf{C}(\mathbf{A}).$$

The $\mathbf{n}\times\mathbf{p}$ matrix $\mathbf{A}^-$ is a one condition generalised inverse, or "g_1-inverse", if and only if

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A};$$

another useful result is that if $\mathbf{B}$ is a $\mathbf{t}\times\mathbf{n}$ matrix,

$$\mathbf{B}\mathbf{A}^-\mathbf{A} = \mathbf{B}$$

if and only if

$$\mathbf{C}(\mathbf{B}') \subseteq \mathbf{C}(\mathbf{A}'): \qquad \langle 1.6.1.1 \rangle$$

see, e.g. Rao and Mitra [1971, lemma 2.2.4.(i)].

Suppose that the $\mathbf{p}\times 1$ vector $\mathbf{x}$ and the $\mathbf{n}\times 1$ vector $\mathbf{z}$ ($\mathbf{p} > \mathbf{n}$) are related by the equation

$$\mathbf{x} = \mathbf{B}\mathbf{z} + \mathbf{c},$$

$\mathbf{B}$ a known $\mathbf{p}\times\mathbf{n}$ matrix of full column rank, and $\mathbf{c}$ a known $\mathbf{p}\times 1$ vector. Let the columns of $\mathbf{B}$ form a basis for $\mathbf{N}(\mathbf{D})$, $\mathbf{D}$ a known $(\mathbf{p}-\mathbf{n})\times\mathbf{p}$ matrix of full row rank. Then,

$$\mathbf{D}\mathbf{x} = \mathbf{D}\mathbf{c} \iff \mathbf{x} = \mathbf{B}\mathbf{z} + \mathbf{c},$$

with

$$\mathbf{C}(\mathbf{B}) = \mathbf{N}(\mathbf{D}).$$

If A is a $p\times n$ matrix, with columns $a_1,\ldots,a_n$, then the $pn\times 1$ vector vec A is defined by

$$(\text{vec } A)' = (a_1',\ldots,a_n')'.$$

If A, B, C and D are matrices such that the products

ABC and $ABCD$

are defined, then

$$\text{vec}(ABC) = (D' \otimes A)\text{vec } B$$

and

$$\text{tr } (ABCD) = (\text{vec } A')'(D' \otimes B)\text{vec } C. \qquad \langle 1.6.1.2\rangle$$

1.6.2. It will turn out to be convenient, from the point of view of obtaining first-order conditions for maximisation, to use certain matrix differential results, which can be found in Magnus and Neudecker [1978] and Neudecker [1969]:

$$d \log \det A = \text{tr } (A^{-1\prime} dA);$$

$$d\, A^{-1} = -A^{-1} dA\, A^{-1};$$

$$d \text{ tr } AB = \text{tr}(dA.B + A.dB);$$

$$\text{vec } dA = d \text{ vec } A;$$

$$d(A \otimes B) = (dA \otimes B) + (A \otimes dB);$$

$$d\, AB = dA.B + A.dB.$$

The first two results require A to be nonsingular, whilst the indicated products are assumed to exist.

1.6.3. Various matrices and vector expressions associated with the commutation matrix K_{pn}, introduced by several authors (see Magnus and Neudecker [1979]) will be found useful in what follows. The $pn\times pn$ commutation matrix K_{pn} has the properties, amongst others, that

$K_{pn} = K'_{pn}$;

if A is $p \times n$, then

$$K_{pn} \operatorname{vec} A = \operatorname{vec} A'; \tag{1.6.3.1}$$

if x is $p \times n$ and y is $n \times 1$, then

$$K_{pn}(y \otimes x) = (x \otimes y); \tag{1.6.3.2}$$

if C is $h \times t$, then

$$K_{hp}(A \otimes C) = (C \otimes A)K_{tn}, \tag{1.6.3.3}$$

or

$$K_{hp}(A \otimes C)K_{nt} = (C \otimes A). \tag{1.6.3.4}$$

The following results are given by Magnus and Neudecker [1978]: denoting K_{nn} by K_n, the $n^2 \times n^2$ matrix S_n is defined as

$$S_n = \tfrac{1}{2}(I_{n^2} + K_n); \tag{1.6.3.5}$$

it is idempotent and symmetric, and

$$S_n K_n = S_n = K_n S_n. \tag{1.6.3.6}$$

If A is $n \times n$, and x is $n \times 1$, then

$$S_n(A \otimes A) = S_n(A \otimes A)S_n = (A \otimes A)S_n; \tag{1.6.3.7}$$

$$S_n(x \otimes x) = (x \otimes x). \tag{1.6.3.8}$$

For the same matrix A, the $\tfrac{1}{2}n(n+1) \times 1$ vector $v(A)$ is the vector obtained from vec A by deleting the ultradiagonal elements of A; if A is symmetric, then $v(A)$ contains the distinct elements of A in a lexicographic ordering. The elimination matrix L_n performs the transformation

$$L_n \operatorname{vec} A = v(A), \tag{1.6.3.9}$$

and hence L_n is $\tfrac{1}{2}n(n+1) \times n^2$; it has the property that

$L_n L'_n = I_{\frac{1}{2}n(n+1)}$.

The matrix $\mathbf{D_n}$ is defined to be the transformation matrix such that if A is $\mathbf{n}$x$\mathbf{n}$ and symmetric,

$$\mathbf{D_n'}v(A) = \text{vec } A; \qquad \langle 1.6.3.10\rangle$$

$\mathbf{D_n}$ has the same dimensions as $\mathbf{L_n}$ and has the following properties:

$$\mathbf{D_n L_n'} = \mathbf{I}_{\frac{1}{2}n(n+1)}; \qquad \langle 1.6.3.11\rangle$$

$$\mathbf{D_n K_n} = \mathbf{D_n} = \mathbf{D_n S_n}; \qquad \langle 1.6.3.12\rangle$$

$$\mathbf{D_n' L_n S_n} = \mathbf{S_n}; \qquad \langle 1.6.3.13\rangle$$

for A $\mathbf{n}$x$\mathbf{n}$,

$$\mathbf{D_n' L_n}(A \otimes A)\mathbf{D_n'} = (A \otimes A)\mathbf{D_n'}. \qquad \langle 1.6.3.14\rangle$$

If A is nonsingular,

$$[\mathbf{D_n}(A \otimes A)\mathbf{D_n'}]^{-1} = \mathbf{L_n S_n}(A^{-1} \otimes A^{-1})\mathbf{S_n L_n'}, \qquad \langle 1.6.3.15\rangle$$

$$[\mathbf{L_n}(A \otimes A)\mathbf{D_n'}]^{-1} = \mathbf{L_n}(A^{-1} \otimes A^{-1})\mathbf{D_n'}. \qquad \langle 1.6.3.16\rangle$$

A slight modification of Theorem 3.3(iv) of Magnus and Neudecker [1978] is required: let A and B be $\mathbf{n}$x$\mathbf{n}$; then

$$\begin{aligned}\mathbf{S_n}[(A \otimes B) + (B \otimes A)] &= \tfrac{1}{2}[\mathbf{I_n}^2 + \mathbf{K_n}][(A \otimes B) + (B \otimes A)] \\ &= \tfrac{1}{2}[(A \otimes B) + (B \otimes A) \\ &\qquad + \mathbf{K_n}(A \otimes B) + \mathbf{K_n}(B \otimes A)] \\ &= \tfrac{1}{2}[(A \otimes B) + (B \otimes A) \\ &\qquad + (B \otimes A)\mathbf{K_n} + (A \otimes B)\mathbf{K_n}] \\ &= [(A \otimes B) + (B \otimes A)]\mathbf{S_n},\end{aligned}$$

and hence

$$\begin{aligned}\mathbf{S_n}[(A \otimes B) + (A \otimes B)]\mathbf{S_n} &= [(A \otimes B) + (B \otimes A)]\mathbf{S_n} \\ &= \mathbf{S_n}[(A \otimes B) + (B \otimes A)].\end{aligned}$$

Furthermore,

$$[(A \otimes B) + (B \otimes A)]\mathbf{S_n} = (A \otimes B)\mathbf{S_n} + \tfrac{1}{2}(B \otimes A) + \tfrac{1}{2}(B \otimes A)\mathbf{K_n}$$

$$= (A \otimes B)S_n + \tfrac{1}{2}K_n(A \otimes B)K_n + \tfrac{1}{2}K_n(A \otimes B)$$

$$= (A \otimes B)S_n + \tfrac{1}{2}K_n(A \otimes B)(I_{n^2} + K_n)$$

$$= (A \otimes B)S_n + K_n(A \otimes B)S_n$$

$$= 2S_n(A \otimes B)S_n$$

$$= 2S_n(B \otimes A)S_n \qquad \langle 1.6.3.17\rangle$$

by a similar argument.

1.6.4. An extension of some of the ideas of subsection 1.6.2. to block diagonal matrices will be required in Chapter 4. Let A be the block diagonal matrix

$$A = \begin{bmatrix} A_1 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & A_h \end{bmatrix},$$

where each A_i is $p_i x p_i$, $i = 1,\ldots, h$. A shorthand notation for A is

$$A = \bigoplus_{i=1}^{h} A_i.$$

To find the vec of such a matrix in terms of vec A_i, $i = 1,\ldots,h$, define the matrix

$$E_i = \begin{bmatrix} 0_{p_1,p_i} \\ \cdot \\ \cdot \\ I_{p_i} \\ \cdot \\ \cdot \\ 0_{p_h,p_i} \end{bmatrix} \qquad \langle 1.6.4.1\rangle$$

of dimension

$(\sum_{i=1}^{h} p_i) \times p_i$;

then,

$$A = (E_1A_1,\dots,E_hA_h),$$

and

$$\begin{aligned}\text{vec } A &= \text{vec}[\text{vec}(E_1A_1),\dots,\text{vec}(E_hA_h)] \\ &= \text{vec}[(I_{p_1} \otimes E_1)\text{vec } A_1,\dots,(I_{p_h} \otimes E_h)\text{vec } A_h] \\ &= \bigoplus_{i=1}^{h}(I_{p_i} \otimes E_i)\begin{bmatrix} \text{vec } A_1 \\ \cdot \\ \cdot \\ \text{vec } A_h \end{bmatrix}. \qquad \langle 1.6.4.2\rangle\end{aligned}$$

One can deduce from this the corresponding result for

$\text{vec}(I_h \otimes A)$,

where A is an $n \times n$ matrix:

$$\text{vec}(I_h \otimes A) = [\bigoplus_{i=1}^{h}(I_n \otimes e_i \otimes I_n)]\begin{bmatrix} I_n \otimes 1 \otimes I_n \\ \cdot \\ \cdot \\ I_n \otimes 1 \otimes I_n \end{bmatrix} \text{vec } A$$

$$= \begin{bmatrix} I_n \otimes e_1 \otimes I_n \\ \cdot \\ \cdot \\ I_n \otimes e_h \otimes I_n \end{bmatrix} \text{vec } A,$$

where e_i, $i = 1,\dots,h$ are h-dimensional coordinate vectors.

If A should be rectangular, $p \times n$, this result becomes

$$\text{vec}(I_h \otimes A) = \begin{bmatrix} I_n \otimes e_1 \otimes I_p \\ \cdot \\ \cdot \\ I_n \otimes e_h \otimes I_p \end{bmatrix} \text{vec } A. \qquad \langle 1.6.4.3\rangle$$

1.6.5. To deal with the distribution of quadratic forms arising from a singular multivariate normal distribution, a special case of Corollary 2.11.1 of Srivastava and Khatri [1979] is extremely useful: if

$$\mathbf{y} \sim N(\mathbf{a}, \mathbf{B}),$$

then $\mathbf{y}'\mathbf{C}\mathbf{y}$ has the non-central χ^2-distribution with $\mathbf{t}$ degrees of freedom, and non-centrality parameter

$$\mathbf{d} = \mathbf{a}'\mathbf{C}\mathbf{a}$$

if and only if

$$\mathrm{tr}\ \mathbf{CB} = \mathrm{rank}\ \mathbf{BCB} = \mathbf{t},$$

$$\mathbf{BCBCB} = \mathbf{BCB},$$

$$\mathbf{BCa} = \mathbf{BCBCa},$$

$$\mathbf{d} = \mathbf{a}'\mathbf{Ca} = \mathbf{a}'\mathbf{BCBa}.$$

The most common application of this result is to the case where $\mathbf{a} = 0$, $\mathbf{C} = \mathbf{B}^-$: if $\mathbf{y} \sim N(0, \mathbf{B})$, then

$$\mathbf{y}'\mathbf{B}^-\mathbf{y} \sim \chi^2_{\mathbf{t}}, \qquad \langle 1.6.5.1 \rangle$$

for

$$\mathbf{BB}^-\mathbf{BB}^-\mathbf{B} = \mathbf{BB}^-\mathbf{B} = \mathbf{B},$$

and

$$\mathrm{tr}(\mathbf{B}^-\mathbf{B}) = \mathrm{rank}\ \mathbf{B}$$

follows from Theorem 1.6.1 of Srivastava and Khatri [1979].

1.6.6. Many of the test statistics described in the succeeding chapters can be calculated from what one might interpret loosely as a generalised least squares (GLS) regression. To make this more precise, let $\mathbf{x}$ be a $\mathbf{p}$-vector, $\mathbf{F}$ a $\mathbf{p}\times\mathbf{n}$ matrix, with $\mathbf{p} > \mathbf{n}$, and $\mathbf{G}$ a $\mathbf{p}\times\mathbf{p}$ positive definite matrix. Then, the process of finding the minimum of

$(\mathbf{x} - \mathbf{F}\mathbf{a})'\mathbf{G}(\mathbf{x} - \mathbf{F}\mathbf{a})$

over $\mathbf{a}$ will be called the "regression of $\mathbf{x}$ on $\mathbf{F}$ in the metric of $\mathbf{G}$", and if $\hat{\mathbf{a}}$ is one of the minimising values, say,

$\hat{\mathbf{a}} = (\mathbf{F}'\mathbf{G}\mathbf{F})^{-}\mathbf{F}'\mathbf{G}\mathbf{x}$,

then

$(\mathbf{x} - \mathbf{F}\hat{\mathbf{a}})'\mathbf{G}(\mathbf{x} - \mathbf{F}\hat{\mathbf{a}}) = \mathbf{x}'[\mathbf{I}_p - \mathbf{G}\mathbf{F}(\mathbf{F}'\mathbf{G}\mathbf{F})^{-}\mathbf{F}']\mathbf{G}[\mathbf{I}_p - \mathbf{F}(\mathbf{F}'\mathbf{G}\mathbf{F})^{-}\mathbf{F}'\mathbf{G}]\mathbf{x}$

will be called the "residual squared norm" (RSN) of the regression. It is also possible to define an "explained squared norm", or ESN, as

$$\begin{aligned}\mathbf{x}'\mathbf{G}\mathbf{x} - \text{RSN} = \text{ESN} \\ &= \mathbf{x}'\mathbf{G}'\mathbf{F}(\mathbf{F}'\mathbf{G}\mathbf{F})^{-}\mathbf{G}'\mathbf{F}\mathbf{x} \\ &= \hat{\mathbf{a}}'\mathbf{F}'\mathbf{G}\mathbf{F}\hat{\mathbf{a}}.\end{aligned}$$

The values of both RSN and ESN can be shown to be invariant to the choice of g_1-inverse, and hence to the choice of solution $\hat{\mathbf{a}}$ of the "normal equations"

$\mathbf{F}'\mathbf{G}\mathbf{F}\mathbf{a} = \mathbf{F}'\mathbf{G}\mathbf{x}$:

see, for example, Seber [1977, p73,81]. Whilst the motivation behind these ideas is that of "curve fitting", in many instances there will be an underlying regression model on which to base interpretations of the regression coefficients.

1.6.7. In the derivation of the limit distributions underlying certain non-nested test statistics, some unusual mean, variance and covariance expressions are required.

Firstly, Theorem 5 of Magnus and Neudecker [1979] states that if the px1 random vector $\mathbf{y}$ is defined by

$\mathbf{y} = \mathbf{c} + \mathbf{u}$,

where $\mathbf{c}$ is the mean of $\mathbf{y}$, and

$\mathbf{u} \sim N(0, \mathbf{M})$,

then, summarising the parameters $\mathbf{c}$, $v(\mathbf{M})$ in the vector κ,

$$E_\kappa \mathrm{vec}(\mathbf{y}\mathbf{y}') = E_\kappa(\mathbf{y} \otimes \mathbf{y}) = \mathrm{vec}\,\mathbf{M} + (\mathbf{c} \otimes \mathbf{c}),$$

and

$$\begin{aligned} \mathrm{var}_\kappa[\mathrm{vec}(\mathbf{y}\mathbf{y}')] &= \mathrm{var}_\kappa(\mathbf{y} \otimes \mathbf{y}) \\ &= 2\mathbf{S_p}[(\mathbf{M} \otimes \mathbf{M}) + (\mathbf{M} \otimes \mathbf{cc}') + (\mathbf{cc}' \otimes \mathbf{M})] \\ &= 2\mathbf{S_p}(\mathbf{M} \otimes \mathbf{M}) + 4\mathbf{S_p}(\mathbf{M} \otimes \mathbf{cc}')\mathbf{S_p}. \end{aligned}$$

Secondly, it will be necessary to find the covariance matrix of a vector having the following structure:

$$\begin{bmatrix} v(\sum_{t=1}^{n} \mathbf{y_t}\mathbf{y_t}') \\ \mathrm{vec}(\sum_{t=1}^{n} \mathbf{x_t}\mathbf{y_t}') \end{bmatrix} = \sum_{t=1}^{n} \begin{bmatrix} v(\mathbf{y_t}\mathbf{y_t}') \\ \mathrm{vec}(\mathbf{x_t}\mathbf{y_t}') \end{bmatrix}$$

where the $\mathbf{p}$x1 random vector $\mathbf{y_t}$ is generated from

$$\mathbf{y_t} = \mathbf{P}'\mathbf{w_t} + \mathbf{u_t}, \quad \mathbf{t} = 1,\ldots,\mathbf{n},$$

$\mathbf{w_t}$ an $\mathbf{f}$x1 and $\mathbf{x_t}$ an $\mathbf{h}$x1 fixed vector, with $\mathbf{u_t}$ independent random vectors, such that

$$\mathbf{u_t} \sim N(0, \mathbf{M}), \quad \mathbf{t} = 1,\ldots,\mathbf{n}.$$

Using equation <1.6.3.9>, the assumed independence, and the results above, one can show that

$$\mathrm{var}_\kappa[\sum_{t=1}^{n} v(\mathbf{y_t}\mathbf{y_t}')] = \mathbf{L_p}[2\mathbf{n}\mathbf{S_p}(\mathbf{M} \otimes \mathbf{M}) + 4\sum_{t=1}^{n}\mathbf{S_p}(\mathbf{M} \otimes \mathbf{P}\mathbf{w_t}\mathbf{w_t}'\mathbf{P}')\mathbf{S_p}]\mathbf{L_p}'$$

(where κ now represents vec $\mathbf{P}$ and $v(\mathbf{M})$), and

$$\mathrm{var}_\kappa[\sum_{t=1}^{n} \mathrm{vec}(\mathbf{x_t}\mathbf{y_t}')] = \sum_{t=1}^{n}(\mathbf{M} \otimes \mathbf{x_t}\mathbf{x_t}').$$

The covariance term is found by the following argument:

$$\begin{aligned} &\mathrm{cov}_\kappa[v(\sum_{t=1}^{n} \mathbf{y_t}\mathbf{y_t}'), \mathrm{vec}(\sum_{t=1}^{n} \mathbf{x_t}\mathbf{y_t}')] \\ &\quad = \sum_{t=1}^{n} \mathrm{cov}_\kappa[v(\mathbf{y_t}\mathbf{y_t}'), \mathrm{vec}(\mathbf{x_t}\mathbf{y_t}')] \end{aligned}$$

$$= L_p \sum_{t=1}^{n} \mathrm{cov}_\kappa [\mathrm{vec}(y_t y_t'), \mathrm{vec}(x_t y_t')]$$

by independence, and

$$\mathrm{cov}_\kappa [\mathrm{vec}(y_t y_t'),\ \mathrm{vec}(x_t y_t')]$$

$$= E_\kappa[(y_t \otimes y_t) - E_\kappa(y_t \otimes y_t)][\mathrm{vec}(x_t u_t')]'$$

$$= E_\kappa[(P'w_t \otimes P'w_t) + (u_t \otimes P'w_t)$$

$$+ (P'w_t \otimes u_t) + (u_t \otimes u_t) - \mathrm{vec}\ M$$

$$- (P'w_t \otimes P'w_t)][u_t \otimes x_t]'$$

$$= E_\kappa[(u_t \otimes P'w_t) + K_p(u_t \otimes P'w_t)$$

$$+ (u_t \otimes u_t) - \mathrm{vec}\ M][u_t' \otimes x_t']$$

$$= 2S_p E_\kappa(u_t u_t' \otimes P'w_t x_t') + E_\kappa(u_t u_t' \otimes u_t x_t')$$

$$= 2S_p(M \otimes P'w_t x_t'),$$

using the fact that the third moments of a multivariate normal distribution are zero.

Hence,

$$\mathrm{cov}_\kappa[v(\sum_{t=1}^{n} y_t y_t'),\ \mathrm{vec}(\sum_{t=1}^{n} x_t y_t')] = 2L_p S_p \sum_{t=1}^{n} (M \otimes P'w_t x_t'),$$

and finally,

$$\mathrm{var}_\kappa \sum_{t=1}^{n} \begin{bmatrix} v(y_t y_t') \\ \mathrm{vec}(x_t y_t') \end{bmatrix}$$

$$= \sum_{t=1}^{n} \begin{bmatrix} 2L_pS_p(M \otimes M)S_pL_p' + 4L_pS_p(M \otimes P'w_t w_t'P)S_pL_p' & \vdots & 2L_pS_p(M \otimes P'w_t x_t') \\ 2(M \otimes x_t w_t'P)S_pL_p' & \vdots & (M \otimes x_t x_t') \end{bmatrix} \qquad \langle 1.6.7.1\rangle$$

A useful special case of this occurs when y_t has mean vector zero: i.e. when $P = 0$; y_t is then identical with u_t. Thus, κ now represents $v(M)$ only, and

$$\mathrm{var}_x \sum_{t=1}^{n} \begin{bmatrix} v(u_t u_t') \\ \mathrm{vec}(x_t u_t') \end{bmatrix} = \sum_{t=1}^{n} \begin{bmatrix} 2L_p S_p (M \otimes M) S_p L_p' & \vdots & 0 \\ 0 & \vdots & (M \otimes x_t x_t') \end{bmatrix}$$

<1.6.7.2>

Appendix to Chapter 1.

The typefaces used in this thesis are set out below, in a comparative form.

Lower Case

Plain : a b c d e f g h i j k l m n o p q r s t u v w x y z

Byte : a b c d e f g h i j k l m n o p q r s t u v w x y z

Bold : **a b c d e f g h i j k l m n o p q r s t u v w x y z**

Outline: **a b c d e f g h i j k l m n o p q r s t u v w x y z**

Upper Case

Plain : A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Byte : A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Bold : **A B C D E F G H I J K L M N O P Q R S T U V W X Y Z**

Outline: **A B C D E F G H I J K L M N O P Q R S T U V W X Y Z**

Italic : *A B C D E F G H I J K L M N O P Q R S T U V W X Y Z*

Chapter 2: Estimation and Inference for Constraint Parameter Problems.

2.1. Introduction.

2.1.1. In this Chapter, it is assumed that each of the independent and observable random vectors $y_1,\ldots,y_n$ has density

$f_t(y_t;\theta)$, $\quad t = 1,\ldots,n$,

so that the log-likelihood function is

$l_n(y;\theta) = \sum_{t=1}^{n}\log f_t(y_t;\theta)$

where

$y' = (y_1',\ldots,y_n')$.

In order to provide a general framework into which the simultaneous equations model may be embedded, the $s_0 \times 1$ parameter θ is assumed to satisfy certain a priori restrictions expressed in "constraint parameter" form. Since the main interest of the chapter is inference, estimation will be discussed under an assumed null hypothesis. The alternative hypothesis, or maintained model, is

$$H_1 : \theta = \phi(\beta), \qquad \langle 2.1.1.1\rangle$$

where β is an $r_1 \times 1$ vector of free or structural parameters. Under the null hypothesis, β satisfies the restrictions

$\beta = \lambda(\alpha)$,

where α is an $r_0 \times 1$ vector of structural parameters. The null hypothesis may be described by

$$H_0 : \theta = \phi(\beta), \quad \beta = \lambda(\alpha), \qquad \langle 2.1.1.2\rangle$$

or equivalently, by

$$H_0 : \theta = \phi[\lambda(\alpha)] = \theta(\alpha). \qquad \langle 2.1.1.3\rangle$$

Under the null hypothesis, the true values of θ and α are denoted θ^0 and α^0 respectively, and in a corresponding way, the true value of β under this hypothesis is

$\beta^0 = \lambda(\alpha^0)$.

The aim of the chapter is to devise tests of the null hypothesis H_0 against the alternative hypothesis H_1, initially using the Likelihood Ratio and Lagrange Multiplier test principles; some problems are encountered in directly constructing a Wald test statistic, which are resolved by using a minimum chi-squared estimator as the basis of such a test statistic. Another type of test statistic, the C-alpha, usually denoted in this thesis by CA, is also defined and discussed.

2.1.2. There are a number of sections in this chapter which act as formal preparation for the discussion of inference: in section 2.2., the almost sure convergence properties of maximum likelihood estimators of θ and β from the alternative hypothesis model <2.1.1.1>, and of θ, α and β from the null hypothesis model <2.1.1.2> are established, assuming the null hypothesis to be true. These results are technically necessary for what follows, and are also useful for the discussion of similar almost sure convergence results in an unidentified model, discussed in Chapter 6.

In section 2.3., it is assumed that the log-likelihood function $l_n(y;\theta)$ possesses continuous derivatives in θ, and that $\phi(\beta)$ and $\lambda(\alpha)$ possess continuous derivatives in their arguments, so the maximum likelihood estimators may be found by finding the turning points of a suitable Lagrangean function. It is shown that under suitable conditions, the estimated Lagrange multipliers $\tilde{\tau}$, $\tilde{\mu}$ associated with the null hypothesis <2.1.1.3> and the alternative hypothesis <2.1.1.1> converge almost surely to zero, when the null hypothesis is true. The nature of the maximum likelihood estimator $\tilde{\phi}$ when the dimension of θ, s_0, equals that of β, r_1, is also investigated: in this case, $\tilde{\phi}$ is the unrestricted maximum likelihood estimator of θ, and will be denoted $\hat{\theta}$.

Given the results of these two sections, a limiting joint normal distribution, under the null hypothesis <2.1.1.3>, is deduced for each of

$n^{1/2}(\tilde{\theta} - \theta^0)$, $\quad n^{1/2}(\tilde{\alpha} - \alpha^0)$, $\quad n^{1/2}\tilde{\tau}$

and

$n^{1/2}(\tilde{\phi} - \theta^0)$, $\quad n^{1/2}(\tilde{\beta} - \beta^0)$, $\quad n^{1/2}\tilde{\mu}$

in section 2.4., whilst in section 2.5., minimum chi-squared estimators are suggested for use in this constraint parameter framework. The first-order conditions for the maximum likelihood estimators derived in section 2.3. and the limit distribution results derived in section 2.4. are used to construct a "two-step" or "linearised maximum likelihood" estimators, and "linearised minimum chi-squared" estimators in sections 2.6. and 2.7. . A good part of the analysis in sections 2.5. - 2.7. consists of verifying that the proposed estimators are asymptotically equivalent to the maximum likelihood estimator.

In sections 2.8. - 2.12., Likelihood Ratio, Lagrange Multiplier, Wald, C-alpha, and other statistics which are obtained as differences of the former statistics are discussed in considerable detail. Quite a lot of attention is given to ways in which the statistics may be calculated, either as by-products of estimation, or by a regression using quantities calculated from the estimation results.

In the final section, the similarities and differences between the various test statistics are evaluated; the statistics are also classified into tests of "specification" and "misspecification".

2.2. Strong Consistency of Constrained Maximum Likelihood Estimators.

2.2.1. To establish that the maximum likelihood estimators $\tilde{\theta}$ and $\tilde{\alpha}$ converge almost surely to the true values $\theta^0 = \theta(\alpha^0)$ and α^0, respectively, under the null hypothesis <2.1.1.3>,

$$H_0 : \theta = \theta(\alpha),$$

requires a number of additional assumptions, which will now be stated. The first major assumption is that the parameter space $\mathbb{A}$, to which α^0 is assumed interior, is compact: this does seem to be an indispensable assumption. Provided that the function $\theta(\alpha)$ is continuous over $\mathbb{A}$, as assumed in subsection 2.1.2., the set

$$\theta(\mathbb{A})$$

is also compact. The consequence of this is that if the sequence of maximisers of

$$n^{-1}l_n(y;\theta)$$

over $\theta(\mathbb{A})$ is $\{\tilde{\theta}_n\}$, then this sequence has at least one limit point.

The second major assumption is that the sequence of continuous functions of θ,

$$n^{-1}l_n(y;\theta)$$

converges almost surely and uniformly in θ to a continuous function

$$l(\theta^0;\theta)$$

of θ, which has a unique maximum at θ^0. The nature of $l(\theta^0;\theta)$, and the critical assumption of a unique maximum at

θ^0 requires some further discussion, which is postponed to the next subsection.

The essence of the first stage of the proof can now be given: let $\{\tilde{\theta}_{n_i}\}$ be an arbitrary subsequence of $\{\tilde{\theta}_n\}$, with limit θ^*; then, for almost all realisations of the random vector y,

$$\left|n_i^{-1}l_{n_i}(y;\tilde{\theta}_{n_i}) - l(\theta^0;\theta^*)\right| \leq \left|n_i^{-1}l_{n_i}(y;\tilde{\theta}_{n_i}) - l(\theta^0;\tilde{\theta}_{n_i})\right| + \left|l(\theta^0;\tilde{\theta}_{n_i}) - l(\theta^0;\theta^*)\right|.$$

By the uniform convergence, there exists an n^0 such that for all $n_i \geq n^0$ and any θ,

$$\left|n_i^{-1}l_{n_i}(y;\tilde{\theta}_{n_i}) - l(\theta^0;\tilde{\theta}_{n_i})\right| < \tfrac{1}{2}\epsilon;$$

by the continuity of $l(\theta^0;\theta)$, there exists an n^* such that for all $n_i \geq n^*$,

$$\left|l(\theta^0;\tilde{\theta}_{n_i}) - l(\theta^0;\theta^*)\right| < \tfrac{1}{2}\epsilon,$$

so that for $n_i \geq \max(n^0,n^*)$,

$$\left|n_i^{-1}l_{n_i}(y;\tilde{\theta}_{n_i}) - l(\theta^0;\theta^*)\right| < \epsilon.$$

That is, for almost all y,

$$n_i^{-1}l_{n_i}(y;\tilde{\theta}_{n_i}) \longrightarrow l(\theta^0;\theta^*),$$

i.e.,

$$n_i^{-1}l_{n_i}(y;\tilde{\theta}_{n_i}) \xrightarrow{a.s.} l(\theta^0;\theta^*).$$

By hypothesis,

$$n_i^{-1}l_{n_i}(y;\tilde{\theta}_{n_i}) \geq n_i^{-1}l_{n_i}(y;\theta^0),$$

and passing to the limit, for almost all y,

$$l(\theta^0;\theta^*) \geq l(\theta^0;\theta^0).$$

Hence, by the unique maximisation assumption,

$$\theta^* = \theta^0.$$

Since the subsequence $\{\tilde{\theta}_{n_i}\}$ was arbitrary, all limit points of the sequence $\{\tilde{\theta}\}$ are equal to θ^*: i.e. to θ^0. That is, for almost all y,

$\tilde{\theta} \rightarrow \theta^0$:

i.e.

$\tilde{\theta} \xrightarrow{a.s.} \theta^0$.

This method of proof has been used by a number of authors, for example, Amemiya [1973], Frydman [1980].

The second stage is to consider the almost sure convergence of $\tilde{\alpha}$: $\tilde{\theta}$ is defined to be the the value $\theta(\tilde{\alpha})$, so that the almost sure limit of $\tilde{\alpha}$, say, α^*, satisfies

$\theta^0 = \theta(\alpha^*)$.

If α^0 is uniquely identified at θ^0, that is, there is a unique solution for α in

$\theta^0 = \theta(\alpha)$,

then, $\alpha^* = \alpha^0$:

i.e.,

$\tilde{\alpha} \xrightarrow{a.s.} \alpha^0$.

Thus, the maximum likelihood estimators of θ and α from the null hypothesis model are strongly consistent when the null hypothesis model <2.1.1.3> is true, given the satisfaction of the assumptions made.

2.2.2. To discuss the almost sure consistency of the maximum likelihood estimators $\tilde{\phi}$ and $\tilde{\beta}$ for θ^0 and $\beta = \lambda(\alpha^0)$ from the alternative hypothesis model described by equation <2.1.1.1>, it is necessary to consider the nature of the almost sure limit of

$n^{-1}l_n(y;\theta)$

under the null hypothesis, that is the function $l(\theta^0;\theta)$ of equation <2.2.1.1>, and the assumption that this function has a unique maximum at $\theta = \theta^0$.

It is convenient to examine the latter assumption first, using the Kullback-Leibler Information Criterion, as defined by Kullback and Leibler [1952] and Kullback [1959] : see also White [1982]. Let $f_t(y_t;\theta)$ and $g_t(y_t;\phi)$ be density functions; then one can define

$$KLIC_t = E_{\theta^0}[\log f_t(y_t;\theta^0) - \log g_t(y_t;\phi)],$$

which is non-negative, and has a global minimum of 0 if and only if

$$f_t(y_t;\theta^0) = g_t(y_t;\phi).$$

Defining the log-likelihood function

$$m_n(y;\phi) = \sum_{t=1}^{n}\log g_t(y_t;\phi),$$

the average Kullback-Leibler Information Criterion, or $AKLIC_n$,

$$\begin{aligned} AKLIC_n &= n^{-1}\sum_{t=1}^{n}KLIC_t \\ &= n^{-1}E_{\theta^0}[l_n(y;\theta) - m_n(y;\phi)] \qquad \langle 2.2.2.1\rangle \end{aligned}$$

is therefore non-negative and has a global minimum of zero if and only if

$$l_n(y;\theta^0) = m_n(y;\phi).$$

This is a more general description than is strictly necessary, but this definition of $AKLIC_n$ can also be applied to the general discussion of test statistics for non-nested hypotheses in Chapter 8.

The application to the problem at hand is straightforward; note that here, $m_n(y;\phi)$ equals $l_n(y;\phi)$. Let the parameter space for β be $\mathcal{B}$, assumed compact, so that under the alternative hypothesis, the θ-parameter space is $\phi(\mathcal{B})$. Under the null hypothesis, β is constrained to lie in the set $\lambda(\hat{A})$ (where $\hat{A}$ is the α-parameter space), so that the set

$$\theta(\hat{A}) = \phi[\lambda(\hat{A})]$$

is a subset of $\phi(\mathcal{B})$. That is,

$$\theta^0 \in \phi(\mathcal{B}).$$

Then, the global minimum of

$$E_{\theta^0}[l_n(y;\theta^0) - l_n(y;\phi)]$$

can occur for ϕ varying over $\phi(\mathcal{B})$, namely, at $\phi = \theta^0$; equivalently, θ^0 maximises

$$E_{\theta^0} l_n(y;\theta)$$

over $\phi(\mathcal{B})$ as well as over $\theta(\hat{A})$.

Implicitly, it has been supposed that $E_{\theta^0} l_n(y;\theta)$ exists, for each θ in some set: if in addition, a Strong Law of Large Numbers holds for $n^{-1}l_n(y;\theta)$, uniformly in θ, one would have, under the null hypothesis,

$$n^{-1}[l_n(y;\theta) - E_{\theta^0} l_n(y;\theta)] \xrightarrow{a.s.} 0,$$

so that if

$$n^{-1}E_{\theta^0} l_n(y;\theta)$$

has a limit, uniformly in θ, it can be identified with the function

$$l(\theta^0;\theta)$$

of subsection 2.1.1. . To then say that $l(\theta^0;\theta)$ has a unique maximum over $\Theta(\hat{A})$ at θ^0 is effectively the same as saying that $E_{\theta^0} l_n(y;\theta)$ has a unique maximum over $\Theta(\hat{A})$ at θ^0, and that this property is preserved in the limit. Indeed, recalling the compactness assumption for $\hat{A}$, if $\{\theta_n\}$ is the sequence of maximisers of

$$n^{-1}E_{\theta^0} l_n(y;\theta)$$

over $\Theta(\hat{A})$, and θ^* a limit point of this sequence, then

$$n^{-1}E_{\theta^0} l_n(y;\theta^*) \geq n^{-1}E_{\theta^0} l_n(y;\theta^0)$$

and

$$\lim_{n\to\infty} E_{\theta^0} l_n(y;\theta^*) \geq \lim_{n\to\infty} E_{\theta^0} l_n(y;\theta^0),$$

or

$$l(\theta^0;\theta^*) \geq l(\theta^0;\theta^0).$$

But, from the properties of the $AKLIC_n$ of equation <2.2.2.1>, $\theta^* = \theta^0$: thus, the demand that $l(\theta^0;\theta)$ has a unique maximum is practically equivalent to demanding the same of $n^{-1}E_{\theta^0} l_n(y;\theta)$. In turn, this latter demand is equivalent to demanding that

$$AKLIC_n = n^{-1}E_{\theta^0}[l_n(y;\theta^0) - l_n(y;\phi)]$$

has a unique global minimum at θ^0. It is worth noting that Bowden [1973] used this property as the basis of a theory of identification which reproduced the results obtained by Rothenberg [1971].

<u>2.2.3.</u> Consider the application of these ideas now to establishing that under the null hypothesis, the maximum likelihood estimators of the parameters of the alternative hypothesis of equation <2.1.1.1>, $\tilde{\Phi}$ and $\tilde{\beta}$ converge almost

surely to θ^0 and

$\beta^0 = \lambda(\alpha^0)$.

The argument of the previous subsection shows that

$E_{\theta^0} l_n(y;\phi)$

has a global maximum over $\phi(\mathbb{B})$ at θ^0, and it is reasonable to assume, in line with the argument of subsection 2.2.1., that $l(\theta^0;\theta)$ has a unique maximum over $\phi(\mathbb{B})$, as well as over $\theta(\mathbb{A})$, at θ^0.

Given this, the argument of subsection 2.2.1. shows that

$\tilde{\phi} \xrightarrow{a.s} \theta^0 = \phi(\beta^0)$;

given that β^0 is uniquely identified at θ^0, it follows that

$\tilde{\beta} \xrightarrow{a.s} \beta^0 = \lambda(\alpha^0)$,

both results being true under the null hypothesis <2.1.1.2> and under the assumptions made.

2.3. First Order Conditions

2.3.1. The preceding section has already assumed that one can find the solutions to the constrained maximum problems

$\max_\theta n^{-1} l_n(y;\theta)$ subject to $\theta = \theta(\alpha)$

and

$\max_\phi n^{-1} l_n(y;\phi)$ subject to $\phi = \phi(\beta)$

whose solutions are the maximum likelihood estimators of the null and alternative hypothesis models respectively. In this section, it is assumed that the method of Lagrange Multipliers can be used to find the global constrained maxima required.

For estimation under the alternative hypothesis <2.1.1.1>,

H_1: $\phi = \phi(\beta)$,

one has to maximise

$$n^{-1} l_n(y;\phi) \text{ subject to } \phi = \phi(\beta); \qquad \langle 2.3.1.1\rangle$$

let $\mathcal{L}_1$ be the Lagrangean for this problem, and μ the associated Lagrange multiplier:

$$\mathcal{L}_1 = n^{-1} l_n(y;\phi) + \mu'(\phi - \phi(\beta)).$$

The first order conditions are

$$\left.\begin{aligned} D_\phi \mathcal{L}_1 &= n^{-1} D_\phi l_n + \mu = 0, \\ D_\beta \mathcal{L}_1 &= -(D_\beta \phi)'\mu \quad = 0 \\ D_\mu \mathcal{L}_1 &= \phi - \phi(\beta) \quad = 0. \end{aligned}\right\} \qquad \langle 2.3.1.2\rangle$$

It is clear from these equations that the estimated Lagrange multiplier, $\tilde{\mu}$, satisfies

$$\tilde{\mu} = -\, n^{-1} D_\phi l_n(y;\tilde{\phi});$$

it will shortly be shown that under the null hypothesis of equation <2.1.1.3>,

$$\tilde{\mu} \xrightarrow{a.s.} 0.$$

It has already been assumed that β is uniquely identified in

$$\phi = \phi(\beta);$$

now consider the case in which the dimensions of ϕ (or θ) and β are the same:

$$s_0 = r_1.$$

Then, it will follow that

$$D_\beta\phi(\beta)$$

is nonsingular, in particular at $\tilde{\beta}$, and hence

$$\tilde{\mu} = 0.$$

Consequently, the maximum likelihood estimator $\tilde{\phi}$ solves

$$n^{-1}D_\phi l_n(y;\tilde{\phi}) = 0,$$

and is therefore the unrestricted maximum likelihood estimator of θ; it will be denoted $\hat{\theta}$. Then, $\tilde{\beta}$ is simply the (unique) solution of the equation

$$\hat{\theta} = \phi(\tilde{\beta}).$$

Using an econometric terminology, this may be described as the "just-identified" case, and β is said to be "just identified". Note that the results of section 2.2. show that this unrestricted estimator, $\hat{\theta}$, converges almost surely to the true value θ^0 under the null hypothesis.

2.3.2. Turning now to estimation under the null hypothesis, it will be convenient to regard the null hypothesis as being defined by equation <2.1.1.2>:

H_0: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$.

Thus, to find the maximum likelihood estimators of θ, β and α under the null hypothesis, it is necessary to maximise

$n^{-1}l_n(y;\theta)$ subject to $\theta = \phi(\beta)$ and $\beta = \lambda(\alpha)$. <2.3.2.1>

Let the Lagrange multipliers be ξ and ζ respectively, with Lagrangean

$$\mathcal{L}_0 = n^{-1}l_n(y;\theta) + \xi'(\theta - \phi(\beta)) + \zeta'(\beta - \lambda(\alpha));$$

the first-order conditions are

$$\left.\begin{aligned} D_\theta\mathcal{L}_0 &= n^{-1}D_\theta l_n + \xi = 0,\\ D_\beta\mathcal{L}_0 &= \zeta - (D_\beta\phi)'\xi = 0,\\ D_\alpha\mathcal{L}_0 &= -(D_\alpha\lambda)'\zeta = 0\\ D_\xi\mathcal{L}_0 &= \theta - \phi(\beta) = 0\\ D_\zeta\mathcal{L}_0 &= \beta - \lambda(\alpha) = 0. \end{aligned}\right\} \qquad \text{<2.3.2.2>}$$

Denote the resultant estimators by

$\tilde{\theta}$, $\tilde{\beta}_0$, $\tilde{\alpha}$, $\tilde{\xi}$, $\tilde{\zeta}$:

note that $\tilde{\beta}_0$ differs from $\tilde{\beta}$, the estimator obtained from the alternative hypothesis model. By eliminating ζ from these first-order conditions, and using the chain rule derivative

$$D_\alpha\theta(\alpha) = D_\beta\phi(\beta)D_\alpha\lambda(\alpha),$$

one can see that the maximum likelihood estimators $\tilde{\theta}$ and $\tilde{\alpha}$ obtained from <2.3.2.2> are the same as those obtained from

$$\max_\theta n^{-1}l_n(y;\theta) \text{ subject to } \theta = \theta(\alpha),$$

and the associated Lagrange multiplier $\tilde{\tau}$ of this problem equals $\tilde{\xi}$.

The Lagrange multipliers $\tilde{\mu}$, $\tilde{\xi}$, $\tilde{\zeta}$ associated with the problems <2.3.1.1> and <2.3.2.1> respectively, can be shown to converge almost surely to zero under the null hypothesis

<2.1.1.3> by the following argument. Both $\tilde{\phi}$ and $\tilde{\theta}$ converge almost surely to θ^0, the true value under the null hypothesis; if it were true that a Strong Law of Large Numbers held for

$n^{-1}D_\theta l_n(y;\theta)$

uniformly in θ, that is,

$n^{-1}[D_\theta l_n - E_{\theta^0}D_\theta l_n] \xrightarrow{a.s.} 0,$

uniformly in θ, then it would follow that

$n^{-1}[D_\theta l_n(y;\tilde{\theta}) - E_{\theta^0}D_\theta l_n(y;\theta^0)] \xrightarrow{a.s.} 0.$

But,

$E_{\theta^0}D_\theta l_n(y;\theta^0) = [\![E_\theta D_\theta l_n]\!]_{\theta^0} = 0,$

and hence, under these assumptions,

$- n^{-1}D_\theta l_n(y;\tilde{\theta}) = \tilde{\xi} \xrightarrow{a.s.} 0.$

It then follows from the first-order conditions <2.3.2.2>, by the assumed continuity of $D_\beta\phi$ (see subsection 2.1.2.) that $\tilde{\zeta} \xrightarrow{a.s.} 0.$

2.4. The Limiting Normal Distribution of the Maximum Likelihood Estimators

2.4.1. In this section, Taylor series expansions of certain terms in the first-order conditions <2.3.1.2> and <2.3.2.2> are used to establish limiting normal distributions for the maximum likelihood estimators $\tilde{\phi}$, $\tilde{\beta}$, $\tilde{\mu}$ and $\tilde{\theta}$, $\tilde{\alpha}$, $\tilde{\xi}$, $\tilde{\zeta}$ under the null hypothesis <2.1.1.3>,

H_0: $\theta = \theta(\alpha)$.

The maximum likelihood estimators of the alternative hypothesis model satisfy

$$\left.\begin{aligned} n^{-1}D_\theta l_n(y;\tilde{\phi}) + \tilde{\mu} &= 0 \\ - D_\beta\phi'(\tilde{\beta})\tilde{\mu} &= 0 \\ \tilde{\phi} - \phi(\tilde{\beta}) &= 0: \end{aligned}\right\} \qquad <2.4.1.1>$$

first-order Taylor series expansions of

$n^{-1}D_\theta l_n(y;\tilde{\phi})$ and $\phi(\tilde{\beta})$

around θ^0 and β^0 respectively yield

$$n^{-1}D_\theta l_n(y;\tilde{\phi}) = n^{-1}D_\theta l_n(y;\theta^0) + n^{-1}D_\theta^2 l_n(y;\bar{\phi})(\tilde{\phi} - \theta^0),$$

$\bar{\phi} \in (\tilde{\phi},\theta^0)$;

$$\phi(\tilde{\beta}) = \phi(\beta^0) + D_\beta\phi(\bar{\beta})(\tilde{\beta} - \beta^0) = \theta^0 + D_\beta\phi(\bar{\beta})(\tilde{\beta} - \beta^0),$$

$\bar{\beta} \in (\tilde{\beta},\beta^0)$,

which can be substituted into <2.4.1.1>. Rearrangement then produces

$$\begin{bmatrix} n^{-1}D_\theta^2 l_n(y;\bar{\Phi}) & : & 0 & : & I_{s_0} \\ 0 & : & 0 & : & -D_\beta\phi'(\tilde{\beta}) \\ I_{s_0} & : & -D_\beta\phi(\bar{\beta}) & : & 0 \end{bmatrix} \begin{bmatrix} n^{1/2}(\tilde{\phi}-\theta^0) \\ n^{1/2}(\tilde{\beta}-\beta^0) \\ n^{1/2}\tilde{\mu} \end{bmatrix} = -\begin{bmatrix} n^{-1/2}D_\theta l_n(y;\theta^0) \\ 0 \\ 0 \end{bmatrix}$$

<2.4.1.2>

Consider first the right hand side vector: under suitable conditions (which will need to be verified in each specific application of the theory),

$n^{-1/2}D_\theta l_n(y;\theta^0)$

will satisfy a Central Limit Theorem. Under conditions which ensure that

$[\![E_{\theta^0}D_\theta l_n]\!]_{\theta^0} = [\![D_\theta E_\theta l_n]\!]_{\theta^0}$,

the vector

$n^{-1/2}D_\theta l_n(y;\theta^0)$

has mean vector zero and a covariance matrix denoted

$$\begin{aligned} I_n(\theta^0) &= n^{-1}\mathrm{var}_{\theta^0}[\![D_\theta l_n]\!] \\ &= n^{-1}[\![E_\theta D_\theta l_n D_\theta l_n']\!]_{\theta^0} \qquad \langle 2.4.1.3\rangle \\ &= -\, n^{-1}[\![E_\theta D_\theta^2 l_n]\!]_{\theta^0}. \end{aligned}$$

Whilst it is more conventional to define the latter expression as "the" "information matrix" of the problem, it will be more convenient in this work to use the expression <2.4.1.3>.

To complete the expression for the limiting normal distribution for $n^{-1/2}D_\theta l_n(y;\theta^0)$, it will be convenient to make the assumption that the matrix function

$I_n(\theta) = n^{-1}[\![E_\theta D_\theta l_n D_\theta l_n'(y;\theta)]\!]$

converges to a limit function $I(\theta)$ uniformly in θ; this assumption is stronger than necessary, but does lead to simplifications in a number of proofs. The usefulness of such an assumption has been shown by Amemiya [1981], White [1982]. A consequence of this assumption is that

$$\mathbf{I}_n(\theta^0) \longrightarrow \mathbf{I}(\theta^0),$$

and hence

$$n^{-1/2} D_\theta l_n(y;\theta^0) \xrightarrow{d} w \sim N(0, \mathbf{I}(\theta^0)). \qquad \langle 2.4.1.4\rangle$$

Consider now the matrix on the left hand side of equation <2.4.1.2>: since $\tilde{\beta} \xrightarrow{a.s.} \beta^0$, it will follow that

$$\bar{\beta} \xrightarrow{a.s.} \beta^0$$

and hence

$$D_\beta \phi(\tilde{\beta}) \xrightarrow{a.s.} D_\beta \phi(\beta^0)$$

$$D_\beta \phi(\bar{\beta}) \xrightarrow{a.s.} D_\beta \phi(\beta^0)$$

by the continuity assumption. Next, it will be necessary to assume that

$$n^{-1} D_\theta^2 l_n(y;\bar{\phi}), \qquad \bar{\phi} \in (\tilde{\phi}, \theta^0)$$

converges to $-\mathbf{I}(\theta^0)$; a common assumption in this case is to demand that the third derivatives of $n^{-1} l_n(y;\theta)$,

$$n^{-1} D^3_{\theta_i,\theta,\theta} l_n(y;\theta)$$

are bounded in probability uniformly in θ. This is a bit messy, and a more elegant, if strong, assumption is that

$$- n^{-1} D_\theta^2 l_n(y;\theta) - \mathbf{I}_n(\theta) \xrightarrow{a.s.} 0 \qquad \langle 2.4.1.5\rangle$$

(elementwise) uniformly in θ. Thus, since $\bar{\phi} \xrightarrow{a.s.} \theta^0$,

$$- n^{-1} D_\theta^2 l_n(y;\bar{\phi}) \xrightarrow{a.s.} \mathbf{I}(\theta^0)$$

as required. Another consequence of this, much used later, is that

$$\mathbf{I}_n(\tilde{\theta}), \quad \mathbf{I}_n(\tilde{\phi}), \quad \mathbf{I}_n(\hat{\theta})$$

all converge almost surely to $\mathbf{I}(\theta^0)$.

Thus, it is true that under the null hypothesis, <2.1.1.3>,

$$\begin{bmatrix} n^{-1}D_\theta^2 l_n(y;\bar\phi): & 0 & : I_{s_0} \\ 0 & : \quad 0 & :-D_\beta\phi'(\bar\beta) \\ I_{s_0} & :-D_\beta\phi(\bar\beta): & 0 \end{bmatrix} \xrightarrow{a.s.} \begin{bmatrix} -\mathbf{I}(\theta^0): & 0 & : I_{s_0} \\ 0 & : \quad 0 & :-\Phi'(\beta^0) \\ I_{s_0} & :-\Phi(\beta^0): & 0 \end{bmatrix}$$

$$= \mathbf{F}_1(\beta^0) \qquad \text{<2.4.1.6>}$$

say, where

$\Phi(\beta) = D_\beta\phi.$

It then follows that

$$n^{1/2}\begin{bmatrix} \tilde\phi - \theta^0 \\ \tilde\beta - \beta^0 \\ \tilde\mu \end{bmatrix} \overset{a}{\approx} -\mathbf{F}_1^{-1}(\beta^0)\begin{bmatrix} I_{s_0} \\ 0 \\ 0 \end{bmatrix} n^{-1/2}D_\theta l_n(y;\theta^0) \qquad \text{<2.4.1.7>}$$

and one can show by laborious partitioned inversion that

$$\mathbf{F}_1^{-1}(\beta^0) = \begin{bmatrix} -\Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi' & : -\Phi(\Phi'\mathbf{I}\Phi)^{-1} & : \quad \mathbf{P}_\Phi \\ -(\Phi'\mathbf{I}\Phi)^{-1}\Phi' & : -(\Phi'\mathbf{I}\Phi)^{-1} & : -(\Phi'\mathbf{I}\Phi)^{-1}\Phi'\mathbf{I} \\ \mathbf{P}_\Phi' & : -\mathbf{I}\Phi(\Phi'\mathbf{I}\Phi)^{-1} & : \quad \mathbf{I}\mathbf{P}_\Phi \end{bmatrix}$$

$$\text{<2.4.1.8>}$$

where the explicit dependence of Φ on β^0 and $\mathbf{I}$ on θ^0 has been deleted for notational convenience. In this equation, $\mathbf{P}_\Phi$ is the projection matrix

$$\begin{aligned} \mathbf{P}_\Phi &= I_{s_0} - \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi'\mathbf{I} \\ &= I_{s_0} - \Phi(\beta^0)[\Phi'(\beta^0)\mathbf{I}(\theta^0)\Phi(\beta^0)]^{-1}\Phi'(\beta^0)\mathbf{I}(\theta^0) \qquad \text{<2.4.1.9>} \\ &= \mathbf{P}_\Phi(\beta^0). \end{aligned}$$

Using the vectors

$$\tilde\psi_1' = (\tilde\phi' : \tilde\beta' : \tilde\mu') \qquad \text{<2.4.1.10>}$$

$$\psi_1^{0\prime} = (\phi^{0\prime} : \beta^{0\prime} : 0') \qquad \text{<2.4.1.11>}$$

as a summary device, it follows that

$n^{1/2}(\tilde\psi_1 - \psi_1^0) \overset{a}{\approx} N(0, \Psi(\tilde\psi_1;\psi_1^0)),$

under the null hypothesis, and where

$$\Psi(\tilde{\psi}_1;\psi_1^0) = \begin{bmatrix} \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi' & : & \Phi(\Phi'\mathbf{I}\Phi)^{-1} & : & 0 \\ (\Phi'\mathbf{I}\Phi)^{-1}\Phi' & : & (\Phi'\mathbf{I}\Phi)^{-1} & : & 0 \\ 0 & : & 0 & : & \mathbf{I}\mathbf{P}_{\Phi} \end{bmatrix}, \qquad \text{<2.4.1.12>}$$

again deleting the arguments in the matrix expressions; $\mathbf{P}_{\Phi}$ is defined in equation <2.4.1.9> above.

<u>2.4.2.</u> In the next section but one, these results will be used to generate some large sample test statistics; for the moment, the same analysis can be used to find the limit distribution of the maximum likelihood estimators $\tilde{\theta}$, $\tilde{\beta}_0$, $\tilde{\alpha}$, $\tilde{\xi}$, $\tilde{\zeta}$ of the parameters of the null hypothesis model <2.1.1.2>, and which satisfy, from equation <2.3.2.2>, the first-order conditions

$$\begin{aligned} n^{-1}D_{\theta}l_n(y;\tilde{\theta}) + \tilde{\xi} &= 0 \\ \tilde{\zeta} - D_{\beta}\phi'(\tilde{\beta}_0)\tilde{\xi} &= 0 \\ - D_{\alpha}\lambda'(\tilde{\alpha})\tilde{\zeta} &= 0 \\ \tilde{\theta} - \phi(\tilde{\beta}_0) &= 0 \\ \tilde{\beta}_0 - \lambda(\tilde{\alpha}) &= 0. \end{aligned}$$

Define

$$\Lambda(\alpha) = D_{\alpha}\lambda;$$

then, first-order Taylor series expansions of $\phi(\tilde{\beta}_0)$, $\lambda(\tilde{\alpha})$ and $n^{-1}D_{\theta}l_n(y;\tilde{\theta})$ around β^0, α^0 and θ^0 will produce a system of equations analogous to equations <2.4.1.2>:

$$\begin{bmatrix} n^{-1}D_\theta^2 l_n(y;\bar\theta) : & 0 : & I_{s_0} : & 0 : & 0 \\ 0 : & 0 : & -\Phi'(\tilde\beta) : & 0 : & I_{r_1} \\ I_{s_0} : & -\Phi(\bar\beta) : & 0 : & 0 : & 0 \\ 0 : & 0 : & 0 : & 0 : & -\Lambda'(\tilde\alpha) \\ 0 : & I_{r_1} : & 0 : & -\Lambda(\tilde\alpha) : & 0 \end{bmatrix} n^{1/2} \begin{bmatrix} \tilde\theta - \theta^0 \\ \tilde\beta_0 - \beta^0 \\ \tilde\xi \\ \tilde\alpha - \alpha^0 \\ \tilde\zeta \end{bmatrix}$$

$$= - \begin{bmatrix} I_{s_0} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} n^{-1/2} D_\theta l_n(y;\theta^0).$$

The reason for this arrangement of the equations will become clear shortly. Making the same arguments as before, one can claim that the matrix on the left converges almost surely to

$$\mathbf{F}_0(\alpha^0) = \begin{bmatrix} -\mathbf{I}(\theta^0) : & 0 : & I_{s_0} : & 0 : & 0 \\ 0 : & 0 : & -\Phi'(\beta^0) : & 0 : & I_{r_1} \\ I_{s_0} : & -\Phi(\beta^0) : & 0 : & 0 : & 0 \\ 0 : & 0 : & 0 : & 0 : & -\Lambda'(\alpha^0) \\ 0 : & I_{r_1} : & 0 : & -\Lambda(\alpha^0) : & 0 \end{bmatrix}.$$

The leading 3x3 block of this matrix is none other than $\mathbf{F}_1(\beta^0)$ of equation <2.4.1.6>, so that its inverse, given in equation <2.4.1.8>, can be used in the partitioned inversion of $\mathbf{F}_0$; in order to conveniently express the submatrices of this inverse, let

$$\mathbf{\Theta}(\alpha) = D_\alpha \theta(\alpha).$$

Again using a vector, ψ_0, as a summary, let

$$\tilde\psi_0' = (\tilde\theta' : \tilde\beta_0' : \tilde\alpha' : \tilde\xi' : \tilde\zeta'), \quad \psi_0^{0\prime} = (\theta^{0\prime} : \beta^{0\prime} : 0' : \alpha^{0\prime} : 0'),$$
<2.4.2.1>

one has

$$n^{1/2}(\tilde{\psi}_0 - \psi_0^0) \overset{a}{\approx} \begin{bmatrix} \Theta[\Theta'I\Theta]^{-1}\Theta' \\ \Lambda[\Theta'I\Theta]^{-1}\Theta' \\ -\mathbf{P}' \\ [\Theta'I\Theta]^{-1}\Theta' \\ -\Phi'\mathbf{P}' \end{bmatrix} n^{-1/2}D_\theta l_n(y;\theta^0), \qquad \langle 2.4.2.2\rangle$$

where $\mathbf{P}$ is the projection matrix

$$\begin{aligned} \mathbf{P} &= \mathbf{I}_{s_0} - \Theta[\Theta'I\Theta]^{-1}\Theta'\mathbf{I} \\ &= \mathbf{I}_{s_0} - \Theta(\alpha^0)[\Theta'(\alpha^0)\mathbf{I}(\theta^0)\Theta(\alpha^0)]^{-1}\Theta'(\alpha^0)\mathbf{I}(\theta^0) \qquad \langle 2.4.2.3\rangle \\ &= \mathbf{P}(\alpha^0). \end{aligned}$$

Note that the dependence of the matrices in <2.4.2.2> on their arguments has been left implicit. The long matrix in <2.4.2.2> is simply the first column block of $\mathbf{F}_0^{-1}(\alpha^0)$, and from this matrix the covariance matrix

$\Psi(\tilde{\psi}_0;\psi_0^0)$

of the limiting normal distribution of

$n^{1/2}(\tilde{\psi}_0 - \psi_0^0)$

can be deduced directly.

Before using these results to provide test statistics, two other estimators will be developed, and which make some use of the analysis above.

2.5. The Minimum Chi-Squared Estimator

2.5.1. When the alternative hypothesis model of <2.1.1.1> is just identified, $\tilde{\phi}$ becomes the unrestricted maximum likelihood estimator $\hat{\theta}$ of θ, obtained simply by maximising $n^{-1}I_n(y;\theta)$. The method to be proposed for obtaining an estimator of α in the null hypothesis model from this unrestricted estimator $\hat{\theta}$ will also provide the basis for the construction of a Wald test statistic for the hypotheses of equations <2.1.1.3> and <2.1.1.1>,

H_0: $\theta = \theta(\alpha)$

H_1: $\theta = \phi(\beta)$,

the latter hypothesis being "unrestricted".

The estimator is the minimum chi-squared estimator (see Rothenberg [1973,pp24-25]), obtained in this situation by minimising

$$n(\hat{\theta} - \theta(\alpha))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha)) \qquad \langle 2.5.1.1\rangle$$

over α: this corresponds to a non-linear regression of $\hat{\theta}$ on $\theta(\alpha)$ with respect to the metric defined by $\mathbf{I}_n(\hat{\theta})$, in the sense defined in subsection 1.6.6. . Denote the resultant estimator of α by α^*, and correspondingly θ^* for θ. How one might numerically compute minimum chi-squared estimates will be discussed later, in section 2.7. . It will now be shown that the large sample properties of α^* and θ^* are the same as those of the maximum likelihood estimators $\tilde{\alpha}$, $\tilde{\theta}$ from the null hypothesis model.

2.5.2. It has been shown in subsections 2.2.3. and 2.3.1. that

$\hat{\theta} \xrightarrow{a.s.} \theta^0$,

$\mathbf{I}_n(\hat{\theta}) \xrightarrow{a.s.} \mathbf{I}(\theta^0)$,

and hence, by continuity,

$(\hat{\theta} - \theta(\alpha))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha)) \xrightarrow{a.s.} (\theta^0 - \theta(\alpha))'\mathbf{I}(\theta^0)(\theta^0 - \theta(\alpha))$;

if, as has been assumed in subsection 2.2.2., α^0 is uniquely identified in

$\theta^0 = \theta(\alpha)$,

then the function

$(\theta^0 - \theta(\alpha))'\mathbf{I}(\theta^0)(\theta^0 - \theta(\alpha))$

will have a unique minimum at α^0, and so by the same type of argument used in subsection 2.2.1.,

$\alpha^\ddagger \xrightarrow{a.s.} \alpha^0$,

$\theta^\ddagger \xrightarrow{a.s.} \theta^0$.

Recall that the parameter space $\mathcal{A}$ for α is assumed to be compact.

2.5.3. To establish that

$n^{1/2}(\theta^\ddagger - \theta^0)$, $\quad n^{1/2}(\alpha^\ddagger - \alpha^0)$

have the same limiting normal distribution as the maximum likelihood estimators $\tilde{\theta}$, $\tilde{\alpha}$, one can combine the Taylor series expansion of $\theta^\ddagger$ around α^0,

$$\alpha^\ddagger = \theta(\alpha^\ddagger) = \theta(\alpha^0) + D_\alpha\theta(\bar{\alpha})(\alpha^\ddagger - \alpha^0) \qquad \text{<2.5.3.1>}$$

$\bar{\alpha} \in (\alpha^\ddagger, \alpha^0)$

with the first-order conditions for the minimisation of <2.5.1.1>,

$$- 2\boldsymbol{\theta}'(\alpha^\ddagger)\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha^\ddagger)) = 0, \qquad \text{<2.5.3.2>}$$

where

$\mathcal{B}(\alpha) = D_\alpha\theta(\alpha)$.

This yields

$$n^{1/2}\mathcal{B}'(\alpha^*)I_n(\hat{\theta})((\hat{\theta} - \theta^0) - \mathcal{B}(\alpha^*)(\alpha^* - \alpha^0)) \xrightarrow{P} 0,$$

or

$$n^{1/2}(\alpha^* - \alpha^0) \overset{a}{\approx} [\mathcal{B}'I\mathcal{B}]^{-1}\mathcal{B}'In^{1/2}(\hat{\theta} - \theta^0): \qquad \langle 2.5.3.3\rangle$$

where the dependence on α^0 and θ^0 has again been suppressed.

From equation <2.4.1.12>, one can deduce that when $\tilde{\phi} = \hat{\theta}$, that is, when the derivative matrix $\Phi(\beta^0)$ is square and non-singular, the covariance matrix of the limiting distribution

$$n^{1/2}(\tilde{\phi} - \theta^0) \overset{a}{\approx} N(0,\Psi(\tilde{\phi};\psi_1^0)), \qquad \langle 2.5.3.4\rangle$$

where ψ_1^0 is defined in equation <2.4.1.11>, collapses from

$$\Psi(\tilde{\phi};\psi_1^0) = \Phi(\Phi'I\Phi)^{-1}\Phi'$$

to

$$\Psi(\hat{\theta};\theta^0) = I^{-1}(\theta^0), \qquad \langle 2.5.3.5\rangle$$

the well known result.

Employing this here leads to the result

$$n^{1/2}(\alpha^* - \alpha^0) \overset{a}{\approx} N(0,\Psi(\alpha^*;\alpha^0)),$$

where

$$\Psi(\alpha^*;\alpha^0) = (\mathcal{B}'I\mathcal{B})^{-1},$$

which is the same as the covariance matrix of the limiting distribution of the maximum likelihood estimator $\tilde{\alpha}$ deduced from equation <2.4.2.2>,

$$n^{1/2}(\tilde{\alpha} - \alpha^0) \overset{a}{\approx} N(0,\Psi(\tilde{\alpha};\alpha^0))$$

with

$$\Psi(\tilde{\alpha};\alpha^0) = (\mathcal{B}'I\mathcal{B})^{-1}.$$

One can go further and show that $n^{1/2}$ times the "residual vector" from the regression of $\hat{\theta}$ on $\theta(\alpha)$ in the metric of $\mathbf{I}_n(\hat{\theta})$ which conceptually produces α^*,

$$n^{1/2}(\hat{\theta} - \theta(\alpha^*))$$

has the same limiting distribution as

$$\mathbf{P}(\alpha^0)\mathbf{I}^{-1}(\theta^0)n^{-1/2}D_\theta l_n(y;\theta^0),$$

where $\mathbf{P}(\alpha^0)$ is defined by equation <2.4.2.3>:

$$\mathbf{P} = \mathbf{I}_{s_0} - \mathbf{\Theta}(\mathbf{\Theta}'\mathbf{I}\mathbf{\Theta})^{-1}\mathbf{\Theta}'\mathbf{I}.$$

For, by the Taylor series expansion <2.5.3.1> of $\theta^* = \theta(\alpha^*)$ and the limit distribution statement <2.5.3.3>,

$$\begin{aligned} n^{1/2}(\hat{\theta} - \theta^*) &\overset{a}{\approx} n^{1/2}(\hat{\theta} - \theta^0) - \mathbf{\Theta}(\alpha^0)n^{1/2}(\alpha^* - \alpha^0) \\ &= [\mathbf{I}_{s_0} - \mathbf{\Theta}(\mathbf{\Theta}'\mathbf{I}\mathbf{\Theta})^{-1}\mathbf{\Theta}'\mathbf{I}]n^{1/2}(\hat{\theta} - \theta^0); \end{aligned}$$

underlying equations <2.5.3.4> and <2.5.3.5> is the assertion

$$n^{1/2}(\hat{\theta} - \theta^0) \overset{a}{\approx} \mathbf{I}^{-1}(\theta^0)n^{-1/2}D_\theta l_n(y;\theta^0).$$

Hence,

$$\begin{aligned} n^{1/2}(\hat{\theta} - \theta^*) &\overset{a}{\approx} \mathbf{P}\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0) \qquad \text{<2.5.3.6>} \\ &= \mathbf{I}^{-1}\mathbf{P}n^{-1/2}D_\theta l_n(y;\theta^0), \end{aligned}$$

as asserted. The usefulness of this result lies in showing that a Wald statistic derived from this minimum chi-squared estimator is asymptotically equivalent to the Likelihood Ratio statistic under the null hypothesis <2.1.1.3>.

<u>2.5.4.</u> Although the discussion of this minimum chi-squared estimator has focussed on how to obtain an estimator of α directly from the unrestricted maximum likelihood estimator $\hat{\theta}$, one can also use the same principle to estimate the parameter α of the null hypothesis model <2.1.1.2>

$$\mathbf{H}_0\text{: } \theta = \phi(\beta), \quad \beta = \lambda(\alpha)$$

using the estimators of θ and β from the alternative hypothesis model <2.1.1.1>

H_1: $\theta = \phi(\beta)$,

namely, $\tilde{\phi}$ and $\tilde{\beta}$. This discussion is only meaningful when $\tilde{\phi} \neq \hat{\theta}$: that is, when $s_0 > r_1$, these being the dimensions of θ and β respectively. From the limiting covariance matrix <2.4.1.12>, one finds

$$n^{1/2}(\tilde{\beta} - \beta^0) \overset{a}{\approx} N(0, \Psi(\tilde{\beta};\beta^0))$$

with

$$\Psi(\tilde{\beta};\beta^0) = (\Phi'(\beta^0)I(\theta^0)\Phi(\beta^0))^{-1}$$

which is consistently estimated by

$$\Psi_n(\tilde{\beta}) = (\Phi'(\tilde{\beta})I_n(\tilde{\theta})\Phi(\tilde{\beta}))^{-1}.$$

Working by analogy with the criterion function <2.5.1.1>, the minimum chi-squared estimator of α, still denoted α^*, is obtained by minimising

$$n[\tilde{\beta} - \lambda(\alpha)]'\Phi'(\tilde{\beta})I_n(\tilde{\phi})\Phi(\tilde{\beta})[\tilde{\beta} - \lambda(\alpha)]$$

with respect to α. This corresponds to a non-linear regression of $\tilde{\beta}$ on $\lambda(\alpha)$ with respect to the metric

$$\Phi'(\tilde{\beta})I_n(\tilde{\phi})\Phi(\tilde{\beta});$$

one can show, by the methods described in subsections 2.5.2. and 2.5.3., that this estimator α^* converges almost surely to α^0, and has the same limit normal distribution as the maximum likelihood estimator $\tilde{\alpha}$, when the null hypothesis is true.

2.6. The Two Step Estimator

2.6.1. The well known Newton-Raphson method is often used for computing the unrestricted maximum likelihood estimator $\hat{\theta}$ of the parameter vector θ; this method stems from a Taylor series expansion of the likelihood equation

$$D_\theta l_n(y;\hat{\theta}) = 0$$

around some value θ^*:

$$0 = D_\theta l_n(y;\hat{\theta}) \simeq D_\theta l_n(y;\theta^*) + D_\theta^2 l_n(y;\theta^*)(\hat{\theta} - \theta^*)$$

implying that

$$\hat{\theta} \simeq \theta^* - (D_\theta^2 l_n(y;\theta^*))^{-1} D_\theta l_n(y;\theta^*),$$

and leading in turn to an iteration scheme

$$\hat{\theta}_{i+1} = \hat{\theta}_i - (D_\theta^2 l_n(y;\hat{\theta}_i))^{-1} D_\theta l_n(y;\hat{\theta}_i).$$

The other aspect of this analysis is that if in fact θ^* is a consistent estimator of θ, such that

$$n^{1/2}(\theta^* - \theta^0) \overset{a}{\approx} N(0,\Psi(\theta^*;\theta^0)),$$

then it is well known (see for example, Harvey [1980, Chapter 4], Zacks [1971, Chapter 5]) that the "two step estimator" $\hat{\hat{\theta}}$ defined by

$$\hat{\hat{\theta}} = \theta^* - (D_\theta^2 l_n(y;\theta^*))^{-1} D_\theta l_n(y;\theta^*)$$

has the same limiting distribution as the unrestricted maximum likelihood estimator $\hat{\theta}$:

$$n^{1/2}(\hat{\hat{\theta}} - \theta^0) \overset{a}{\approx} N(0,I^{-1}(\theta^0)).$$

The version of the Newton-Raphson method and of the two-step estimator which replaces

$$- D_\theta^2 l_n(y;\theta)$$

by its expected value is called the "method of scoring" (see for example, Rao [1973, p370]) and will be employed almost exclusively in the discussions of two-step estimators; the reason for this is that the structure of the resultant two-step estimator is much more amenable to analysis. Denoting this estimator still by $\hat{\theta}$, it is defined as

$$\hat{\theta} = \theta^* + n^{-1}I_n^{-1}(\theta^*)D_\theta l_n(y;\theta^*). \qquad \langle 2.6.1.1\rangle$$

2.6.2. Aitchison and Silvey [1958] extended this idea to the case where θ is subject to constraint equations of the form

$$\mathbf{h}(\theta) = 0,$$

and it can be carried over to the general constraint parameter case described in section 2.1., in particular the null hypothesis model of equation <2.1.1.3>,

$$\mathbf{H}_0\text{: } \theta = \theta(\alpha).$$

One simply uses a Taylor series expansion of the first-order conditions for the problem

$$\max_\theta n^{-1}l_n(y;\theta) \text{ subject to } \theta = \theta(\alpha),$$

which are the same as those given in equation <2.3.1.2>, except for the change of model (and hence of notation):

$$\left.\begin{aligned} n^{-1}D_\theta l_n + \tau &= 0 \\ -(D_\alpha\theta)'\tau &= 0 \\ \theta - \theta(\alpha) &= 0; \end{aligned}\right\} \qquad \langle 2.6.2.1\rangle$$

here, τ is the Lagrange multiplier associated with the constrained maximum problem.

Let θ^* and α^* denote estimators of θ and α such that

$$\theta^* \xrightarrow{P} \theta^0, \quad \alpha^* \xrightarrow{P} \alpha^0$$

and there is a joint limiting normal distribution for

$n^{1/2}(\theta^* - \theta^0), \quad n^{1/2}(\alpha^* - \alpha^0)$:

the reason for these demands will become clear shortly. The two-step estimators of θ, α and τ are then obtained by replacing $n^{-1}D_\theta l_n$ and $\theta(\alpha)$ in the first-order conditions by the zero and first-order terms in their Taylor series expansions about θ^*, α^*,

$$n^{-1}D_\theta l_n \simeq n^{-1}D_\theta l_n(y;\theta^*) + n^{-1}D_\theta^2 l_n(y;\theta^*)(\theta - \theta^*)$$

$$\theta(\alpha) \simeq \theta(\alpha^*) + D_\alpha\theta(\alpha^*)(\alpha - \alpha^*),$$

and solving for the unknowns θ, α, and τ. In line with the discussion of the "method of scoring" in subsection 2.6.1., the second derivative matrix

$$n^{-1}D_\theta^2 l_n(y;\theta^*)$$

will be replaced by the negative of the corresponding estimator of the finite sample information matrix $\mathbf{I}_n(\theta^0)$, i.e. $-\mathbf{I}_n(\theta^*)$.

The solutions $\hat{\theta}$, $\hat{\alpha}$, $\hat{\tau}$ of the resultant system of equations are the required two-step estimators of θ, α, and τ:

$$\begin{bmatrix} -\mathbf{I}_n(\theta^*) & : & 0 & : & I_{s_0} \\ 0 & : & 0 & : & -\mathbf{\Theta}'(\alpha^*) \\ I_{s_0} & : & -\mathbf{\Theta}(\alpha^*) & : & 0 \end{bmatrix} \begin{bmatrix} \hat{\theta}-\theta^* \\ \hat{\alpha}-\alpha^* \\ \hat{\tau} \end{bmatrix} = -\begin{bmatrix} I_{s_0} \\ 0 \\ 0 \end{bmatrix} n^{-1}D_\theta l_n(y;\theta^*).$$

<2.6.2.2>

Defining the left hand matrix to be $\mathbf{F}_n(\alpha^*)$, it has the same structure as the matrix $\mathbf{F}_1$ defined in equation <2.4.1.6>, so that its inverse, equation <2.4.1.8>, can be used to find the solution as

$$\begin{bmatrix} \hat{\theta}-\theta^* \\ \hat{\alpha}-\alpha^* \\ \hat{\tau} \end{bmatrix} = -\begin{bmatrix} -B(\alpha^*)(B'(\alpha^*)I_n(\theta^*)B(\alpha^*))^{-1}B'(\alpha^*) \\ -(B'(\alpha^*)I_n(\theta^*)B(\alpha^*))^{-1}B'(\alpha^*) \\ P_n'(\alpha^*) \end{bmatrix} n^{-1}D_\theta l_n(y;\theta^*),$$

<2.6.2.3>

where

$$P_n(\alpha^*) = I_{s_0} - B(\alpha^*)(B'(\alpha^*)I_n(\theta^*)B(\alpha^*))^{-1}B'(\alpha^*)I_n(\theta^*)$$

<2.6.2.4>

is the finite sample analogue of the projection matrix $\mathbf{P}$ of equation <2.4.2.3>.

If one were to regard this solution as the basis of an iterative method for finding the maximum likelihood estimates $\tilde{\theta}$, $\tilde{\alpha}$, $\tilde{\tau}$, from the null hypothesis model, the iteration for α would be

$$\hat{\alpha}_{i+1} = \hat{\alpha}_i + (B'(\hat{\alpha}_i)I_n(\hat{\theta}_i)B(\hat{\alpha}_i))^{-1}B'(\hat{\alpha}_i)I_n(\hat{\theta}_i)n^{-1}D_\theta l_n(y;\hat{\theta}_i)$$

<2.6.2.5>

which has the interesting feature of using the estimated "score" vector $n^{-1}D_\theta l_n(y;\hat{\theta}_i)$ directly in the update term.

<u>2.6.3.</u> Given these results, the assumptions that

$$\alpha^* \xrightarrow{P} \alpha^0, \quad I_n(\theta^*) \xrightarrow{P} I(\theta^0)$$

and that $n^{1/2}(\alpha^* - \alpha^0)$ has a limit normal distribution under the null hypothesis, it is straightforward to establish that $\hat{\theta}$, $\hat{\alpha}$, $\hat{\tau}$, have the same limit normal distribution as the maximum likelihood estimators $\tilde{\theta}$, $\tilde{\alpha}$, $\tilde{\tau}$. Without this property, two-step estimators would not be so useful in inference.

The matrix $F_n(\alpha^*)$ converges in probability to the matrix

$$F(\alpha^0) = \begin{bmatrix} -I(\theta^0) & : & 0 & : & I_{s_0} \\ 0 & : & 0 & : & -B'(\alpha^0) \\ I_{s_0} & : & -B(\alpha^0) & : & 0 \end{bmatrix}, \qquad \text{<2.6.3.1>}$$

whilst, using a Taylor series expansion of $n^{-\frac{1}{2}}D_\theta l_n(y;\theta^*)$ around θ^0,

$$n^{-\frac{1}{2}}D_\theta l_n(y;\theta^*) = n^{-\frac{1}{2}}D_\theta l_n(y;\theta^0) + n^{-1}D_\theta^2 l_n(y;\bar{\theta})n^{\frac{1}{2}}(\theta^* - \theta^0),$$

(where $\bar{\theta} \in (\theta^*, \theta^0)$), it follows that

$$n^{-\frac{1}{2}}D_\theta l_n(y;\theta^*) \overset{a}{\approx} w - I(\theta^0)n^{\frac{1}{2}}(\theta^* - \theta^0), \qquad \text{<2.6.3.2>}$$

where

$$n^{-\frac{1}{2}}D_\theta l_n(y;\theta^0) \xrightarrow{d} w \sim N(0, I(\theta^0)).$$

Putting all this together, and defining

$$\psi^{*\prime} = (\theta^{*\prime}, \alpha^{*\prime}, 0'), \qquad \hat{\psi}' = (\hat{\theta}', \hat{\alpha}', \hat{\tau}'),$$

$$\psi^{0\prime} = (\theta^{0\prime}, \alpha^{0\prime}, 0'),$$

one has, from equations <2.6.2.3>, <2.6.3.1> and <2.6.3.2>,

$$n^{\frac{1}{2}}(\hat{\psi} - \psi^*) \overset{a}{\approx} -F^{-1}(\alpha^0)\begin{bmatrix} I_{s_0} \\ 0 \\ 0 \end{bmatrix}(w - I(\theta^0)n^{\frac{1}{2}}(\theta^* - \theta^0)):$$

i.e.

$$n^{\frac{1}{2}}(\hat{\psi} - \psi^0) \overset{a}{\approx} n^{\frac{1}{2}}(\psi^* - \psi^0) - F^{-1}(\alpha^0)\begin{bmatrix} I_{s_0} \\ 0 \\ 0 \end{bmatrix}(w - I(\theta^0)n^{\frac{1}{2}}(\theta^* - \theta^0)).$$

Consider the terms

$$n^{\frac{1}{2}}(\psi^* - \psi^0) + F^{-1}(\alpha^0)\begin{bmatrix} I_{s_0} \\ 0 \\ 0 \end{bmatrix}I(\theta^0)n^{\frac{1}{2}}(\theta^* - \theta^0):$$

since

$$n^{\frac{1}{2}}(\theta^* - \theta^0) \overset{a}{\approx} B(\alpha^0)n^{\frac{1}{2}}(\alpha^* - \alpha^0),$$

this expression equals (on deleting the explicit dependence of $\mathbf{P}$, B and $\mathbf{I}$ on α^0 and θ^0)

$$\begin{bmatrix} n^{\frac{1}{2}}(\theta^* - \theta^0) - B[B'\mathbf{I}B]^{-1}B'\mathbf{I}Bn^{\frac{1}{2}}(\alpha^* - \alpha^0) \\ n^{\frac{1}{2}}(\alpha^* - \alpha^0) - \quad [B'\mathbf{I}B]^{-1}B'\mathbf{I}Bn^{\frac{1}{2}}(\alpha^* - \alpha^0) \\ - \mathbf{P}'\mathbf{I}Bn^{\frac{1}{2}}(\alpha^* - \alpha^0) \end{bmatrix} \xrightarrow{P} 0,$$

for,

$$\mathbf{P}'\mathbf{I}B = \mathbf{I}B - \mathbf{I}B[B'\mathbf{I}B]^{-1}B'\mathbf{I}B = 0.$$

Hence,

$$n^{\frac{1}{2}}(\hat{\psi} - \psi^0) \underset{\sim}{\overset{a}{=}} - \mathbf{F}^{-1}(\alpha^0) \begin{bmatrix} \mathbf{I}_{s_0} \\ 0 \\ 0 \end{bmatrix} \mathbf{w}$$

$$= - \begin{bmatrix} - B[B'\mathbf{I}B]^{-1}B' \\ - [B'\mathbf{I}B]^{-1}B' \\ - \mathbf{P}' \end{bmatrix} n^{-\frac{1}{2}}D_\theta l_n(y;\theta^0)$$

$$\underset{\sim}{\overset{a}{=}} n^{\frac{1}{2}} \begin{bmatrix} \tilde{\theta} - \theta^0 \\ \tilde{\alpha} - \alpha^0 \\ \tilde{\tau} \end{bmatrix}$$

from equation <2.4.2.2>.

2.6.4. Thus, the two-step estimators of θ and α in the null hypothesis model <2.1.1.3> have the same limiting distribution as the maximum likelihood estimators $\tilde{\theta}$, $\tilde{\alpha}$, and it will turn out that tests of the hypothesis <2.1.1.3> against <2.1.1.1> using such two-step estimators have the same asymptotic properties as tests based on the maximum likelihood estimators.

This analysis can be extended directly to cover estimation of the null hypothesis model when the relationship between θ and α is expressed by <2.1.1.2> as

$\mathbf{H}_0$: $\quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha),$

as in subsection 2.4.2.: letting $\hat{\psi}_0$ and ψ_0^* now denote the two-step and the initial estimator respectively,

$\hat{\psi}_0' = [\, \hat{\theta}', \hat{\beta}_0', \hat{\xi}', \hat{\alpha}', \hat{\zeta}'],\qquad \psi_0^{*\prime} = [\, \theta^{*\prime}, \beta^{*\prime}, 0', \alpha^{*\prime}, 0']$

analogously to the definitions of the maximum likelihood estimator $\tilde{\psi}_0$ and the vector of true values ψ_0^0 in equation <2.4.2.1>, one can show in exactly the same way that $\hat{\psi}_0$ is generated by

$$\hat{\psi}_0 = \psi_0^* + \begin{bmatrix} \mathrm{B}(\alpha^*)[\mathrm{B}'(\alpha^*)\mathbf{I}_n(\theta^*)\mathrm{B}(\alpha^*)]^{-1}\mathrm{B}'(\alpha^*) \\ \Lambda(\alpha^*)[\mathrm{B}'(\alpha^*)\mathbf{I}_n(\theta^*)\mathrm{B}(\alpha^*)]^{-1}\mathrm{B}'(\alpha^*) \\ -\mathbf{P}_n'(\alpha^*) \\ [\mathrm{B}'(\alpha^*)\mathbf{I}_n(\theta^*)\mathrm{B}(\alpha^*)]^{-1}\mathrm{B}'(\alpha^*) \\ -\Phi'(\beta^*)\mathbf{P}_n'(\alpha^*) \end{bmatrix} n^{-1}D_\theta l_n(y;\theta^*), \qquad \langle 2.6.4.1\rangle$$

where the matrix $\mathbf{P}_n(\alpha)$ is defined in equation <2.6.2.4> above. Similarly, it can be established that this two-step estimator has the same limiting distribution as the maximum likelihood estimator $\tilde{\psi}_0$ under the null hypothesis <2.1.1.2>:

$n^{1/2}(\hat{\psi}_0 - \psi_0^0) \overset{a}{\sim} n^{1/2}(\tilde{\psi}_0 - \psi_0^0).$

This will be useful in constructing tests of the two hypotheses <2.1.1.1> and <2.1.1.3> based on the initial estimator ψ_0^* rather than the two-step estimator $\hat{\psi}_0$.

2.7. A Linearised Minimum Chi-squared Estimator.

2.7.1. In the analysis of two-step estimation, a critical role is played by

$$n^{-1}D_{\theta}l_n(y;\theta^*):$$

clearly, problems will occur if θ^* is actually equal to $\hat{\theta}$, the unrestricted maximum likelihood estimator, since

$$D_{\theta}l_n(y;\hat{\theta}) \equiv 0.$$

In contrast, in subsection 2.5.1., it was noted that the minimum chi-squared estimator of α under the null hypothesis

<2.1.1.3>

$$\mathbf{H}_0\text{:}\ \theta = \theta(\alpha)$$

requires a non-linear regression of the unrestricted maximum likelihood estimator $\hat{\theta}$ on the function $\theta(\alpha)$ in the metric of $\mathbf{I}_n(\hat{\theta})$: that is, the minimum chi-squared estimator, $\alpha^{\ddagger}$ is found by minimising <2.5.1.1>,

$$n(\hat{\theta} - \theta(\alpha))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha)).$$

The first-order conditions are, from <2.5.3.2>,

$$-\ 2\boldsymbol{\theta}'(\alpha^{\ddagger})\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha^{\ddagger})) = 0;$$

using a Taylor series expansion of $\theta(\alpha)$ around the point α^* corresponding to the initial estimators θ^* and α^* such that

$$\theta^* = \theta(\alpha^*),$$

$$\theta(\alpha) \simeq \theta(\alpha^*) + \boldsymbol{\theta}(\alpha^*)(\alpha - \alpha^*)$$

in the same way as in the construction of the two-step estimator above, the linearised minimum chi-squared estimator of α, denoted α^{∇}, is the solution of

$$\boldsymbol{\theta}'(\alpha^*)\mathbf{I}_n(\hat{\theta})[(\hat{\theta} - \theta^*) - \boldsymbol{\theta}(\alpha^*)(\alpha - \alpha^*)] = 0$$

i.e.,

$$\alpha^{\nabla} = \alpha^{*} + [\mathrm{B}'(\alpha^{*})\mathbf{I}_n(\hat{\theta})\mathrm{B}(\alpha^{*})]^{-1}\mathrm{B}'(\alpha^{*})\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta^{*}). \qquad \langle 2.7.1.1\rangle$$

If one iterates on α^{*} and θ^{*} in this expression (but not on terms involving $\hat{\theta}$), the convergent iterate will yield the minimum chi-squared estimators $\theta^{\ddagger}$, $\alpha^{\ddagger}$. It is easy to see that this linearised minimum chi-squared estimator α^{∇} minimises

$$[(\hat{\theta}-\theta^{*}) - \mathrm{B}(\alpha^{*})(\alpha^{\nabla}-\alpha^{*})]'\mathbf{I}_n(\hat{\theta})[(\hat{\theta}-\theta^{*}) - \mathrm{B}(\alpha^{*})(\alpha^{\nabla}-\alpha^{*})], \qquad \langle 2.7.1.2\rangle$$

and corresponds to a regression of $\hat{\theta} - \theta^{*}$ on the regressor matrix $\mathrm{B}(\alpha^{*})$ in the metric of $\mathbf{I}_n(\hat{\theta})$.

Providing that the initial estimators θ^{*} and α^{*} are consistent and have proper limit normal distributions under the null hypothesis, the linearised minimum chi-squared estimator α^{∇} will have the same asymptotic properties as the minimum chi-squared estimator $\alpha^{\ddagger}$.

<u>2.7.2.</u> It is interesting to compare the iteration scheme based on equation <2.7.1.1>, where $\mathbf{I}_n(\hat{\theta})$ is replaced by $\mathbf{I}_n(\theta^{*})$; with updating of $\mathbf{I}_n$,

$$\alpha^{\nabla}_{i+1} = \alpha^{\nabla}_{i} + [\mathrm{B}'(\alpha^{\nabla}_{i})\mathbf{I}_n(\theta^{\nabla}_{i})\mathrm{B}(\alpha^{\nabla}_{i})]^{-1}\mathrm{B}'(\alpha^{\nabla}_{i})\mathbf{I}_n(\theta^{\nabla}_{i})(\hat{\theta} - \theta^{\nabla}_{i}),$$

whilst the updating scheme <2.6.2.5> for the two-step estimator of α, under the same circumstances, is given by

$$\hat{\alpha}_{i+1} = \hat{\alpha}_{i} + [\mathrm{B}'(\hat{\alpha}_{i})\mathbf{I}_n(\hat{\theta}_{i})\mathrm{B}(\hat{\alpha}_{i})]^{-1}\mathrm{B}'(\hat{\alpha}_{i})n^{-1}D_{\theta}l_n(y;\hat{\theta}_{i}).$$

Ignoring the fact that $\hat{\theta}_i$ and θ^{∇}_i may well have different values, one can see that

$n^{-1}D_{\theta}l_n(y;\hat{\theta}_i)$

and

$\mathbf{I}_n(\theta_1^{\nabla})(\hat{\theta} - \theta_1^{\nabla})$

are approximations to each other: compare equation <2.6.1.1>.

These distinctions may seem to be minor, and of little practical importance, but are of interest in the construction of Wald test statistics later on in this chapter.

2.7.3. Another linearised minimum chi-squared estimator can be obtained, for the case in which the null hypothesis <2.1.1.3> is written as

$\mathbf{H}_0$: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$

with the alternative hypothesis being equation <2.1.1.1>,

$\mathbf{H}_1$: $\theta = \phi(\beta)$.

The maximum likelihood estimators of the latter model are denoted $\tilde{\phi}$, $\tilde{\beta}$, as in subsection 2.5.4. . As before, inefficient estimators of the null hypothesis model, α^*, θ^* and

$\beta^* = \lambda(\alpha^*)$

are assumed to be available, and a linearised minimum chi-squared estimator α_0^{∇} arises from the minimisation of

$[(\tilde{\beta}-\beta^*)-\Lambda(\alpha^*)(\alpha_0^{\nabla}-\alpha^*)]'\Phi'(\tilde{\beta})\mathbf{I}_n(\tilde{\phi})\Phi(\tilde{\beta})[(\tilde{\beta}-\beta^*)-\Lambda(\alpha^*)(\alpha_0^{\nabla}-\alpha^*)]$,

where

$\Lambda(\alpha) = D_\alpha\lambda(\alpha)$.

This estimator also has an obvious regression interpretation; its inclusion in this discussion is slightly superfluous, since in the intended application, the linear simultaneous equations model, the function $\lambda(\alpha)$ is linear in α, so that linearisation is unnecessary.

2.8. The Likelihood Ratio Statistic.

In the rest of this chapter, a variety of test statistics for testing the hypotheses <2.1.1.1> and <2.1.1.2>,

$\mathbf{H}_0$: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$,

$\mathbf{H}_1$: $\theta = \phi(\beta)$

are discussed, and their asymptotic equivalence under the null hypothesis shown. Behaviour of the test statistics is investigated in the special case where $\mathbf{H}_1$ is just identified in the sense that the dimensions of θ and β are the same, and further where this hypothesis can be written explicitly as

$\mathbf{H}_2$: θ is unrestricted,

so that the just-identifying information that $\theta = \phi(\beta)$ is not required.

2.8.1. Under the hypotheses <2.1.1.2> and <2.1.1.1>, the respective estimators are $\tilde{\theta}$, $\tilde{\beta}_0$, $\tilde{\alpha}$ and $\tilde{\phi}$, $\tilde{\beta}$, so that the Likelihood Ratio test statistic is

$$LR = -2[l_n(y;\tilde{\theta}) - l_n(y;\tilde{\phi})]. \qquad \text{<2.8.1.1>}$$

It is of some interest for what follows to verify that this statistic has a limiting χ^2-distribution with degrees of freedom equal to the difference of the dimensions of β and α,

$\mathbf{r}_1 - \mathbf{r}_0$,

when the null hypothesis is true.

The simplest way of showing this is to consider second-order Taylor series expansions of $l_n(y;\tilde{\theta})$ and $l_n(y;\tilde{\phi})$ around the unrestricted maximum likelihood estimator $\hat{\theta}$:

$$l_n(y;\tilde{\theta}) = l_n(y;\hat{\theta}) + D_\theta l_n'(y;\hat{\theta})(\tilde{\theta} - \hat{\theta})$$
$$+ \tfrac{1}{2}(\tilde{\theta} - \hat{\theta})' D_\theta^2 l_n(y;\bar{\theta})(\tilde{\theta} - \hat{\theta});$$
$$l_n(y;\tilde{\phi}) = l_n(y;\hat{\theta}) + D_\theta l_n'(y;\hat{\theta})(\tilde{\phi} - \hat{\theta})$$
$$+ \tfrac{1}{2}(\tilde{\phi} - \hat{\theta})' D_\theta^2 l_n(y;\bar{\phi})(\tilde{\phi} - \hat{\theta}),$$

where $\bar{\theta} \in (\tilde{\theta},\hat{\theta})$, $\bar{\phi} \in (\tilde{\phi},\hat{\theta})$. Using the assumption made in subsection 2.4.1., equation <2.4.1.5>, that

$$- n^{-1}D_\theta^2 l_n(y;\theta) - \mathbf{I}_n(\theta) \xrightarrow{a.s.} 0,$$

uniformly in θ, it follows that both

$$n^{-1}D_\theta^2 l_n(y;\bar{\theta}), \quad n^{-1}D_\theta^2 l_n(y;\bar{\phi})$$

converge almost surely to $- \mathbf{I}(\theta^0)$.

Thus, combining the two expansions,

$$LR = -2[l_n(y;\tilde{\theta}) - l_n(y;\tilde{\phi})] \stackrel{a}{\approx} n(\tilde{\theta} - \hat{\theta})'\mathbf{I}(\theta^0)(\tilde{\theta} - \hat{\theta})$$
$$- n(\tilde{\phi} - \hat{\theta})'\mathbf{I}(\theta^0)(\tilde{\phi} - \hat{\theta})$$
$$= n(\tilde{\theta} - \tilde{\phi})'\mathbf{I}(\theta^0)(\tilde{\theta} - \tilde{\phi})$$

<2.8.1.2>

which will now be shown.

From equations <2.4.1.6> and <2.4.1.7>, one can deduce the well known fact that

$$n^{1/2}(\hat{\theta} - \theta^0) \stackrel{a}{\approx} \mathbf{I}^{-1}(\theta^0)n^{-1/2}D_\theta l_n(y;\theta^0) \qquad \text{<2.8.1.3>}$$
$$\stackrel{a}{\approx} N(0,\mathbf{I}^{-1}(\theta^0)),$$

by allowing, as in subsection 2.5.3., the derivative matrix

$$D_\beta\phi(\beta) = \Phi(\beta)$$

to be square and non-singular. Next, from equation <2.4.2.2>, one has

$$n^{1/2}(\tilde{\theta} - \theta^0) \stackrel{a}{\approx} B(\alpha^0)[B'(\alpha^0)\mathbf{I}(\theta^0)B(\alpha^0)]^{-1}B'(\alpha^0)n^{-1/2}D_\theta l_n(y;\theta^0),$$

so that (ignoring the α and θ dependence of B and $\mathbf{I}$)

$$n^{1/2}(\tilde{\theta} - \hat{\theta}) \stackrel{a}{\approx} - [\mathbf{I}^{-1} - \Theta(\Theta'\mathbf{I}\Theta)^{-1}\Theta']n^{-1/2}D_\theta l_n(y;\theta^0)$$

$$\stackrel{a}{\approx} - \mathbf{P}\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0),$$

where $\mathbf{P}$ is defined by equation <2.4.2.3>:

$$\mathbf{P} = \mathbf{I}_{s_0} - \Theta(\Theta'\mathbf{I}\Theta)^{-1}\Theta'\mathbf{I}.$$

Similarly, combining equations <2.4.1.10>, <2.4.1.7> and <2.8.1.3>,

$$n^{1/2}(\tilde{\phi} - \hat{\theta}) \stackrel{a}{\approx} - [\mathbf{I}^{-1} - \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi']n^{-1/2}D_\theta l_n(y;\theta^0)$$

$$\stackrel{a}{\approx} - \mathbf{P}_\Phi\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0),$$

with $\mathbf{P}_\Phi$ defined in equation <2.4.1.9.> as

$$\mathbf{P}_\Phi = \mathbf{I}_{s_0} - \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi'\mathbf{I}.$$

These results can now be employed in the quadratic forms appearing in equation <2.8.1.2> above:

$$n(\tilde{\theta} - \hat{\theta})'\mathbf{I}(\theta^0)(\tilde{\theta} - \hat{\theta}) \stackrel{a}{\approx} n^{-1/2}D_\theta l_n'(y;\theta^0)\mathbf{I}^{-1}\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0)$$

and

$$n(\tilde{\phi} - \hat{\theta})'\mathbf{I}(\theta^0)(\tilde{\phi} - \hat{\theta}) \stackrel{a}{\approx} n^{-1/2}D_\theta l_n'(y;\theta^0)\mathbf{I}^{-1}\mathbf{P}_\Phi'\mathbf{I}\mathbf{P}_\Phi\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0);$$

the matrices $\mathbf{P}$ and $\mathbf{P}_\Phi$ share the properties that

$$\mathbf{P}'\mathbf{I} = \mathbf{I}\mathbf{P}, \quad \mathbf{P}_\Phi'\mathbf{I} = \mathbf{I}\mathbf{P}_\Phi;$$

$$\mathbf{P}\mathbf{I}^{-1} = \mathbf{I}^{-1}\mathbf{P}', \quad \mathbf{P}_\Phi\mathbf{I}^{-1} = \mathbf{I}^{-1}\mathbf{P}_\Phi'.$$

Consequently, one can write the limiting distribution of the right hand side of equation <2.8.1.2> as

$$n^{-1/2}D_\theta l_n'(y;\theta^0)(\mathbf{P} - \mathbf{P}_\Phi)\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0).$$

It then follows from the expressions above for $n^{1/2}(\tilde{\theta}-\hat{\theta})$ and $n^{1/2}(\tilde{\phi}-\hat{\theta})$ that

$$n^{1/2}(\tilde{\theta} - \tilde{\phi}) \stackrel{a}{\approx} - (\mathbf{P} - \mathbf{P}_\Phi)\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0)$$

and hence

$$n(\tilde{\theta} - \tilde{\phi})'\mathbf{I}(\tilde{\theta} - \tilde{\phi}) \stackrel{a}{\approx} n^{-1/2}D_\theta l_n'(y;\theta^0)(\mathbf{P} - \mathbf{P}_\Phi)\mathbf{I}^{-1}n^{-1/2}D_\theta l_n(y;\theta^0)$$

$$\stackrel{a}{\approx} n^{-1/2}D_\theta l_n'(y;\theta^0)(\mathbf{P} - \mathbf{P}_\Phi)\mathbf{I}^{-1}$$

$$\times \; (\mathbf{P} - \mathbf{P}_{\Phi})' n^{-1/2} D_{\theta} l_n(y;\theta^0).$$

To show that the LR statistic has the asserted limit χ^2-distribution, let

$$n^{-1/2} D_{\theta} l_n(y;\theta^0) \xrightarrow{d} \mathbf{w} \sim N(0, \mathbf{I}(\theta^0)),$$

as in subsection 2.4.1., equation <2.4.1.4>. Then,

$$LR \xrightarrow{d} \mathbf{w}'(\mathbf{P} - \mathbf{P}_{\Phi})\mathbf{I}^{-1}(\mathbf{P} - \mathbf{P}_{\Phi})'\mathbf{w}: \qquad \langle 2.8.1.4\rangle$$

one can show that the matrix

$$(\mathbf{P} - \mathbf{P}_{\Phi})\mathbf{I}^{-1}(\mathbf{P} - \mathbf{P}_{\Phi})'$$

satisfies the conditions (given in subsection 1.6.5.) required for the quadratic form in $\mathbf{w}$ to have a central χ^2-distribution. For, using the notation of subsection 1.6.5.,

$$\mathbf{B} = \mathbf{I}, \quad \mathbf{C} = (\mathbf{P} - \mathbf{P}_{\Phi})\mathbf{I}^{-1}(\mathbf{P} - \mathbf{P}_{\Phi})',$$

it follows that

$$\mathbf{CBC} = (\mathbf{P} - \mathbf{P}_{\Phi})\mathbf{I}^{-1}(\mathbf{P} - \mathbf{P}_{\Phi})'\mathbf{I}(\mathbf{P} - \mathbf{P}_{\Phi})\mathbf{I}^{-1}(\mathbf{P} - \mathbf{P}_{\Phi})'$$
$$= \mathbf{C},$$

so that

$$\mathbf{BCBCB} = \mathbf{BCB}$$

holds, and

$$\mathrm{tr}\;\mathbf{CB} = \mathrm{tr}\;(\mathbf{P} - \mathbf{P}_{\Phi}) = (s_0 - r_0) - (s_0 - r_1) = r_1 - r_0.$$

Thus,

$$LR \overset{a}{\sim} \chi^2_{r_1 - r_0}$$

under the null hypothesis <2.1.1.3>.

2.8.2. In the special case where the alternative hypothesis <2.1.1.1>,

$$H_1: \quad \theta = \phi(\beta)$$

is just identified, $\tilde{\Phi}$ coincides with the unrestricted estimator $\hat{\theta}$, and the LR statistic is simply

$$LR = -2[l_n(y;\tilde{\theta}) - l_n(y;\hat{\theta})]$$

$$\overset{a}{\sim} \chi^2_{r_1 - r_0},$$

by the analysis of the previous subsection. Here, however, the dimension of β, r_1, equals that of θ, s_0, so that the degrees of freedom amount to $s_0 - r_0$.

Identical results are obtained for the case where θ is completely unrestricted under the hypothesis

H_2: $\theta \neq \theta(\alpha)$,

or θ unrestricted, since the maximum value of $l_n(y;\theta)$ under H_2 is $l_n(y;\hat{\theta})$.

2.9. The Lagrange Multiplier Statistic.

2.9.1. The intuitive basis of this statistic, due to Aitchison and Silvey [1958], Silvey [1959], is very simple: when the null hypothesis of equation <2.1.1.3>,

$\mathbf{H}_0$: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$

is true, the additional restrictions

$\beta = \lambda(\alpha)$

imposed on the alternative hypothesis <2.1.1.1>

$\mathbf{H}_1$: $\theta = \phi(\beta)$

should have an associated Lagrange multiplier which is zero. In the derivation of maximum likelihood estimators under the null hypothesis of subsection 2.3.2., the estimated Lagrange multipliers are $\tilde{\xi}$, (for $\theta = \phi(\beta)$) and $\tilde{\zeta}$ (for $\beta = \lambda(\alpha)$), so that a natural test can be based on the limiting distribution of $n^{1/2}\tilde{\zeta}$, which, from the limiting distribution summary of equation <2.4.2.2>, is

$$n^{1/2}\tilde{\zeta} \overset{a}{\approx} -\Phi'(\beta^0)\mathbf{P}'(\alpha^0)n^{-1/2}D_\theta \mathbf{l}_n(y;\theta^0)$$

$$\overset{a}{\approx} N(0,\Phi'\mathbf{P}'\mathbf{I}\mathbf{P}\Phi),$$

where $\mathbf{P}$ is defined by <2.4.2.3>:

$$\mathbf{P} = \mathbf{I}_{s_0} - \mathbf{\Theta}[\mathbf{\Theta}'\mathbf{I}\mathbf{\Theta}]^{-1}\mathbf{\Theta}'\mathbf{I}$$

(with deletion of α and θ dependence again). It is also true, from the first-order conditions <2.3.2.2>, that

$$\tilde{\zeta} - \Phi'(\tilde{\beta}_0)\tilde{\xi} = 0,$$

whilst

$$-\Lambda'(\tilde{\alpha})\tilde{\xi} = 0,$$

so that the limiting distribution of $n^{1/2}\tilde{\zeta}$ can be interpreted as arising from

$n^{1/2}\tilde{\xi} \sim N(0, \mathbf{P}'\mathbf{I}\mathbf{P})$,

from the distribution summary <2.4.2.2>.

The Lagrange Multiplier test statistic is then

$$LM = n\tilde{\zeta}'[\Phi'(\tilde{\beta}_0)\mathbf{P}'(\tilde{\alpha})\mathbf{I}_n(\tilde{\theta})\mathbf{P}(\tilde{\alpha})\Phi(\tilde{\beta}_0)]^{-}\tilde{\zeta},$$

and it will now be shown that this statistic has the same limiting distribution as the Likelihood Ratio statistic under the null hypothesis. The basic method for doing this is to find a particularly natural g-inverse for

$$\Phi'\mathbf{P}'\mathbf{I}\mathbf{P}\Phi.$$

Introduce the projection

$$\begin{aligned}\mathbf{P}_\Lambda(\alpha) = \mathbf{P}_\Lambda &= \mathbf{I}_{r_1} - \Lambda(\Lambda'\Phi'\mathbf{I}\Phi\Lambda)^{-1}\Lambda'\Phi'\mathbf{I}\Phi \\ &= \mathbf{I}_{r_1} - \Lambda(\boldsymbol{\Theta}'\mathbf{I}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}'\mathbf{I}\Phi,\end{aligned}$$

since

$$\begin{aligned}\boldsymbol{\Theta} = D_\alpha\theta &= D_\beta\phi(\beta)D_\alpha\lambda(\alpha) \\ &= \Phi(\beta)\Lambda(\alpha);\end{aligned}$$

$\mathbf{P}_\Lambda$ has the property that

$$\Phi\mathbf{P}_\Lambda = \mathbf{P}\Phi,$$

and then

$$\Phi'\mathbf{P}'\mathbf{I}\mathbf{P}\Phi = \mathbf{P}_\Lambda'\Phi'\mathbf{I}\Phi\mathbf{P}_\Lambda,$$

which has the easily verified g-inverse

$$\mathbf{P}_\Lambda(\Phi'\mathbf{I}\Phi)^{-1}\mathbf{P}_\Lambda'.$$

Using this in the LM statistic above, one obtains

$$LM = n\tilde{\zeta}'\mathbf{P}_\Lambda(\tilde{\alpha})[\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi'(\tilde{\beta}_0)]^{-1}\mathbf{P}_\Lambda'(\tilde{\alpha})\tilde{\zeta}; \qquad \text{<2.9.1.1>}$$

however, from the first-order conditions <2.3.2.2>,

$$\begin{aligned}\mathbf{P}_\Lambda'(\tilde{\alpha})\tilde{\zeta} &= \tilde{\zeta} - \Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\boldsymbol{\Theta}(\tilde{\alpha})[\boldsymbol{\Theta}'(\tilde{\alpha})\mathbf{I}_n(\tilde{\theta})\boldsymbol{\Theta}(\tilde{\alpha})]^{-1}\Lambda'(\tilde{\alpha})\tilde{\zeta} \\ &= \tilde{\zeta}\end{aligned}$$

$$= \Phi'(\tilde{\beta}_0)\tilde{\xi},$$

which yields in turn

$$LM = n\tilde{\xi}'\Phi(\tilde{\beta}_0)[\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi(\tilde{\beta}_0)]^{-1}\Phi'(\tilde{\beta}_0)\tilde{\xi}. \qquad \text{<2.9.1.2>}$$

From the fact (see equation <2.4.2.2>) that

$$n^{\frac{1}{2}}\tilde{\xi} \overset{a}{\approx} - \mathbf{P}'(\alpha^0)n^{-\frac{1}{2}}D_\theta l_n(y;\theta^0),$$

it can be seen that

$$LM \overset{a}{\approx} n^{-1}D_\theta l_n'(y;\theta^0)\mathbf{P}\Phi[\Phi'\mathbf{I}\Phi]^{-1}\Phi'\mathbf{P}'D_\theta l_n(y;\theta^0),$$

whilst using the property that

$$\Theta = \Phi\Lambda$$

yields

$$\mathbf{P}\Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi'\mathbf{P}' = \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi' - \Theta(\Theta'\mathbf{I}\Theta)^{-1}\Theta'$$

$$= (\mathbf{P} - \mathbf{P}_\Phi)\mathbf{I}^{-1},$$

with $\mathbf{P}_\Phi$ as defined by <2.4.1.9>:

$$\mathbf{P}_\Phi = \mathbf{I}_{s_0} - \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi'\mathbf{I}.$$

Thus, using the limiting random vector $\mathbf{w}$ such that

$$n^{-\frac{1}{2}}D_\theta l_n(y;\theta^0) \xrightarrow{d} \mathbf{w},$$

$$LM \overset{a}{\approx} \mathbf{w}'(\mathbf{P} - \mathbf{P}_\Phi)\mathbf{I}^{-1}(\mathbf{P} - \mathbf{P}_\Phi)'\mathbf{w}, \qquad \text{<2.9.1.3>}$$

$$\overset{a}{\approx} LR,$$

comparing equation <2.8.1.4>.

2.9.2. It is more usual to work with the "score statistic" version of the Lagrange Multiplier statistic (see Rao [1948], Breusch and Pagan [1980]) which makes use of the fact, from the first-order conditions <2.3.2.2>,

$$\tilde{\xi} = - n^{-1}D_\theta l_n(y;\tilde{\theta});$$

when this is substituted in equation <2.9.1.2>, one obtains

$$LM = n^{-1}D_\theta l_n'(y;\tilde{\theta})\Phi(\tilde{\beta}_0)[\Phi'(\tilde{\beta}_0)I_n(\tilde{\theta})\Phi(\tilde{\beta}_0)]^{-1}\Phi'(\tilde{\beta}_0)D_\theta l_n(y;\tilde{\theta}). \qquad \langle 2.9.2.1\rangle$$

This can be regarded as n times the "explained squared norm" from the regression of

$n^{-1}I_n^{-1}D_\theta l_n(y;\tilde{\theta})$

on $\Phi(\tilde{\beta}_0)$ in the metric of $I_n(\tilde{\theta})$; see subsection 1.6.6. for an explanation of "explained squared norm".

2.9.3. In the special case where the alternative hypothesis <2.1.1.1>,

H_1: $\theta = \phi(\beta)$

is just identified, the matrices $\Phi(\tilde{\beta}_0)$ and $\Phi(\beta^0)$ are nonsingular, and hence the Lagrange Multiplier statistic of equation <2.9.2.1> collapses to

$$LM = n^{-1}D_\theta l_n'(y;\tilde{\theta})I_n^{-1}(\tilde{\theta})D_\theta l_n(y;\tilde{\theta}) \qquad \langle 2.9.3.1\rangle$$
$$= n\tilde{\xi}'I_n^{-1}(\tilde{\theta})\tilde{\xi}.$$

If one were to obtain the estimators $\tilde{\theta}$, $\tilde{\alpha}$ under the null hypothesis <2.1.1.3>,

H_0: $\theta = \theta(\alpha)$

simply by maximising

$n^{-1}l_n(y;\theta)$ subject to $\theta = \theta(\alpha)$,

as in subsection 2.6.2. (in connection with two-step estimation), the Lagrange Multiplier statistic for a test against the alternative hypothesis

H_2: θ unrestricted,

or equivalently,

H_2: $\theta \neq \theta(\alpha)$,

would coincide with the statistic <2.9.3.1> above.

This displays the fact that the structure of the model $\phi(\beta)$ under a just identified alternative hypothesis is irrelevant to the value of a Lagrange Multiplier statistic: this leads to certain problems of interpretation of the results of such a test. A proper discussion of these issues is postponed to section 2.13., so that all the various test statistics can be considered together.

2.10. Wald Test Statistics.

2.10.1 Conventional discussions of Wald test statistics in this likelihood context would typically suppose that the parameter vector θ is constrained by the vector equation

$$\mathbf{g}(\theta) = 0,$$

and one can solve this equation to be in the form used in this thesis,

$$\theta = \theta(\alpha)$$

in general only locally, say around the true value θ^0. A conversion of these latter equations into $\mathbf{g}(\theta) = 0$ is also possible, locally.

It has been argued in subsection 1.3.2. that the simultaneous equations model is embedded most naturally into the constraint parameter formulation discussed in this chapter, so that it is sensible to try to construct a Wald test statistic for testing the hypotheses <2.1.1.2> and <2.1.1.1>,

$$\mathbf{H}_0: \quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha)$$

$$\mathbf{H}_1: \quad \theta = \phi(\beta)$$

without resorting to the conversion of these "freedom equations" into "constraint equations".

2.10.2. Consider the case where the alternative hypothesis $\mathbf{H}_1$ is just identified: then, a constraint equation Wald test statistic would examine whether $\mathbf{g}(\hat{\theta})$ is "small"; a corresponding argument for the constraint parameter case might be to examine whether

$\hat{\theta} - \theta(\alpha^*)$

is small, where α^* is the minimum chi-squared estimator of α, discussed in section 2.5. .

Recall from equation <2.5.1.1> that α^* is found as the solution to the problem

$\min_\alpha \, (\hat{\theta} - \theta(\alpha))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha)),$

that is, it is found from the non-linear regression of $\hat{\theta}$ on $\theta(\alpha)$ in the metric of $\mathbf{I}_n(\hat{\theta})$, and that, from subsection 2.5.3., equation <2.5.3.6>,

$n^{1/2}(\hat{\theta} - \theta(\alpha^*)) \overset{a}{\approx} \mathbf{I}^{-1}(\theta^0)\mathbf{P}'(\alpha^0)n^{-1/2}D_\theta l_n(y;\theta^0)$

under the null hypothesis <2.1.1.3>, where $\mathbf{P}(\alpha^0)$ is defined by equation <2.4.2.3>:

$\mathbf{P}(\alpha^0) = \mathbf{I}_{s_0} - \mathbf{B}(\alpha^0)[\mathbf{B}'(\alpha^0)\mathbf{I}(\theta^0)\mathbf{B}(\alpha^0)]^{-1}\mathbf{B}'(\alpha^0)\mathbf{I}(\theta^0).$

It follows from this that

$n^{1/2}(\hat{\theta} - \theta(\alpha^*)) \overset{a}{\approx} N(0, \mathbf{I}^{-1}\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{I}^{-1})$

so that

$n(\hat{\theta} - \theta(\alpha^*))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha^*)) \overset{a}{\approx} \chi^2_{r_1-r_0},$

$\mathbf{r}_1$ being the dimension of β (and hence of θ in this case), and $\mathbf{r}_0$ the dimension of α. To verify this limiting distribution assertion, note that

$n(\hat{\theta}-\theta(\alpha^*))'\mathbf{I}_n(\hat{\theta})(\hat{\theta}-\theta(\alpha^*)) - n(\hat{\theta}-\theta(\alpha^*))'\mathbf{I}(\theta^0)(\hat{\theta}-\theta(\alpha^*)) \xrightarrow{P} 0,$

and that a suitable g-inverse for $\mathbf{I}^{-1}\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{I}^{-1}$ is $\mathbf{I}(\theta^0)$:

$\mathbf{I}^{-1}\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{I}^{-1}\mathbf{I}\mathbf{P}\mathbf{I}^{-1}\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{I}^{-1} = \mathbf{I}^{-1}\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{P}\mathbf{I}^{-1} = \mathbf{I}^{-1}\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{I}^{-1}$

as required.

That is, a suitable Wald test statistic is given by n times the residual squared norm,

$$W = n(\hat{\theta} - \theta(\alpha^*))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha^*)) \qquad \text{<2.10.2.1>}$$

associated with the minimum chi-squared estimator α^*. One can verify that this statistic is asymptotically equivalent to the Lagrange Multiplier test statistic, given in this case by equation <2.9.3.1>:

$$LM = n\tilde{\xi}'\mathbf{I}_n^{-1}(\tilde{\theta})\tilde{\xi}.$$

For, from the limiting distribution summary of equation <2.4.2.2>,

$$n^{½}\tilde{\xi} \overset{a}{\approx} - \mathbf{P}'(\alpha^0)n^{-½}D_\theta \mathbf{l}_n(y;\theta^0),$$

whilst, as noted above,

$$n^{½}(\hat{\theta} - \theta(\alpha^*)) \overset{a}{\approx} - \mathbf{I}^{-1}(\theta^0)\mathbf{P}'(\alpha^0)n^{-½}D_\theta \mathbf{l}_n(y;\theta^0).$$

This minimum chi-squared based test statistic has proved useful in inference problems in demand systems, where the unrestricted model is easy to estimate, and there are a few, rather complex, restrictions.

2.10.3. When the alternative hypothesis is overidentified, the minimum chi-squared estimator of α uses the relationship

$$\beta = \lambda(\alpha)$$

between the free parameters in the null and alternative hypotheses,

$$\mathbf{H}_0\text{:} \quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha)$$

$$\mathbf{H}_1\text{:} \quad \theta = \phi(\beta),$$

that is, α^* now minimises

$$(\tilde{\beta} - \lambda(\alpha))'\Phi'(\tilde{\beta})\mathbf{I}_n(\tilde{\phi})\Phi(\tilde{\beta})(\tilde{\beta} - \lambda(\alpha)),$$

and the appropriate test statistic is n times the minimum of this expression, that is, n times the residual squared norm of the non-linear regression of $\tilde{\beta}$ on $\lambda(\alpha)$ in the metric of $\Phi'(\tilde{\beta})I_n(\tilde{\phi})\Phi(\tilde{\beta})$.

One can show directly that this test statistic too is asymptotically equivalent to the appropriate Lagrange Multiplier test statistic, namely, <2.9.2.1>: from <2.4.1.7> and <2.4.1.8>, it follows that

$$n^{1/2}(\tilde{\beta} - \beta^0) \overset{a}{\approx} [\Phi'(\beta^0)I(\theta^0)\Phi(\beta^0)]^{-1}\Phi'(\beta^0)n^{-1/2}D_\theta l_n(y;\theta^0)$$

whilst for the estimator of β implied by α^*,

$$\beta^* = \lambda(\alpha^*),$$

$$n^{1/2}(\beta^* - \beta^0) \overset{a}{\approx} \Lambda(\alpha^0)[\mathrm{B}'(\alpha^0)I(\theta^0)\mathrm{B}(\alpha^0)]^{-1}\mathrm{B}'(\alpha^0)n^{-1/2}D_\theta l_n(y;\theta^0),$$

from the fact that the minimum chi-squared estimator has the same limiting distribution as the maximum likelihood estimator:

$$n^{1/2}(\alpha^* - \alpha^0) \overset{a}{\approx} n^{1/2}(\tilde{\alpha} - \alpha^0).$$

One can write

$$W \overset{a}{\approx} n(\tilde{\beta} - \beta^*)'\Phi'(\beta^0)I(\theta^0)I^{-1}(\theta^0)I(\theta^0)\Phi(\beta^0)(\tilde{\beta} - \beta^*),$$

and

$$\begin{aligned} I(\theta^0)\Phi(\beta^0)n^{1/2}(\tilde{\beta}-\beta^*) &\overset{a}{\approx} [I\Phi(\Phi'I\Phi)^{-1}\Phi'-I\Phi\Lambda(\mathrm{B}'I\mathrm{B})^{-1}\mathrm{B}']n^{-1/2}D_\theta l_n(y;\theta^0) \\ &\overset{a}{\approx} (\mathbf{P} - \mathbf{P}_\Phi)'n^{-1/2}D_\theta l_n(y;\theta^0) \end{aligned}$$

since

$$\mathrm{B}(\alpha^0) = \Phi(\beta^0)\Lambda(\alpha^0);$$

$\mathbf{P}$ is defined by equation <2.4.2.3> and $\mathbf{P}_\Phi$ by equation <2.4.1.9>. Since

$$(\mathbf{P} - \mathbf{P}_\Phi)I^{-1}(\mathbf{P} - \mathbf{P}_\Phi)' = (\mathbf{P} - \mathbf{P}_\Phi)I^{-1},$$

it is clear that this Wald test statistic is asymptotically equivalent to the Lagrange Multiplier and Likelihood Ratio test statistics. The degrees of freedom of the limit χ^2-distribution under the null hypothesis are $\mathbf{r}_1 - \mathbf{r}_0$, as before, except that in this case, $\mathbf{s}_0 \neq \mathbf{r}_1$.

2.10.4. It is quite easy now to obtain the corresponding Wald test statistic from the linearised minimum chi-squared estimators of section 2.7., which arise from the linear regression of $\hat{\theta} - \theta^*$ on $\Theta(\alpha^*)$ with respect to the metric $\mathbf{I}_n(\hat{\theta})$ in the just identified case, and of $\tilde{\beta} - \beta^*$ on $\Lambda(\alpha^*)$ in the metric of $\Phi'(\tilde{\beta})\mathbf{I}_n(\tilde{\phi})\Phi(\tilde{\beta})$ for the over-identified case. The test statistics are simply n times the residual squared norms from these linear regressions:

$$n[(\hat{\theta}-\theta^*)-\Theta(\alpha^*)(\alpha^{\nabla}-\alpha^*)]'\mathbf{I}_n(\hat{\theta})[(\hat{\theta}-\theta^*)-\Theta(\alpha^*)(\alpha^{\nabla}-\alpha^*)]$$

and

$$n[(\tilde{\beta}-\beta^*)-\Lambda(\alpha^*)(\alpha_0^{\nabla}-\alpha^*)]'\Phi'(\tilde{\beta})\mathbf{I}_n(\tilde{\phi})\Phi(\tilde{\beta})[(\tilde{\beta}-\beta^*)-\Lambda(\alpha^*)(\alpha_0^{\nabla}-\alpha^*)],$$

where α^*, β^*, θ^* are the appropriate initial estimators, and α^{∇}, α_0^{∇} the corresponding linearised minimum chi-squared estimators of α. That these statistics have the correct limiting χ^2-distributions follows from the fact that the minimum chi-squared estimator

$$n^{1/2}(\alpha^{\ddagger} - \alpha^0)$$

and the linearised minimum chi-squared estimators

$$n^{1/2}(\alpha^{\nabla} - \alpha^0), \quad n^{1/2}(\alpha_0^{\nabla} - \alpha^0)$$

have the same limiting distributions.

2.11. The C-alpha Statistic.

2.11.1. The two-step estimator

$$\hat{\psi}'_0 = (\hat{\theta}', \hat{\beta}'_0, \hat{\xi}', \hat{\alpha}', \hat{\zeta}')$$

defined in equation <2.6.4.1> corresponds to estimation of the null hypothesis model of equation <2.1.1.2> in the form

$$\mathbf{H}_0: \quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha).$$

It was shown in section 2.6. that this estimator has the same limiting normal distribution (under the null hypothesis) as the corresponding maximum likelihood estimator

$$\tilde{\psi}'_0 = (\tilde{\theta}', \tilde{\beta}'_0, \tilde{\xi}', \tilde{\alpha}', \tilde{\zeta}');$$

one could consider using this two-step estimator directly to form an analogue of the Lagrange Multiplier statistic.

The Lagrange Multiplier statistic can be given in three equivalent forms, equation <2.9.1.1>, which is less useful for the current purpose, the expression <2.9.1.2>

$$LM = n\tilde{\xi}'\Phi(\tilde{\beta}_0)[\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi(\tilde{\beta}_0)]^{-1}\Phi'(\tilde{\beta}_0)\tilde{\xi},$$

and the "score" version, equation <2.9.2.1>,

$$LM = n^{-1}D_\theta\mathbf{l}'_n(y;\tilde{\theta})\Phi(\tilde{\beta}_0)[\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi(\tilde{\beta}_0)]^{-1}\Phi'(\tilde{\beta}_0)D_\theta\mathbf{l}_n(y;\tilde{\theta}).$$

The analogue of the LM statistic arises essentially from the replacement of $\tilde{\xi}$ by $\hat{\xi}$: from equation <2.6.4.1>,

$$\hat{\xi} = -\mathbf{P}'_n(\alpha^*)n^{-1}D_\theta\mathbf{l}_n(y:\theta^*),$$

where, from equation <2.6.2.4>,

$$\mathbf{P}_n(\alpha^*) = \mathbf{I}_{s_0} - \Theta(\alpha^*)[\Theta'(\alpha^*)\mathbf{I}_n(\theta^*)\Theta(\alpha^*)]^{-1}\Theta'(\alpha^*)\mathbf{I}_n(\theta^*),$$

and

$$\theta^* = \theta(\alpha^*).$$

By evaluating $\Phi(\beta)$ at

$\beta^* = \theta(\alpha^*)$

and $\mathbf{I}_n(\theta)$ at θ^*, a statistic expressed wholly in terms of the initial inefficient estimators is obtained: this is the "C-alpha" statistic,

$$CA = n^{-1}D_\theta \mathbf{l}_n'(y;\theta^*)\mathbf{P}_n(\alpha^*)\Phi(\beta^*)[\Phi'(\beta^*)\mathbf{I}_n(\theta^*)\Phi(\beta^*)]^{-1} \times \Phi'(\beta^*)\mathbf{P}_n'(\alpha^*)D_\theta \mathbf{l}_n(y;\theta^*). \qquad \langle 2.11.1.1\rangle$$

That this has the same limit $\chi^2_{r_1-r_0}$ distribution as the Lagrange Multiplier statistic follows from the asymptotic equivalence of the maximum likelihood and two-step estimators, provided that the initial estimators α^*, β^*, θ^* satisfy the conditions of consistency and asymptotic normality stated in subsection 2.6.2. .

The C-alpha test statistic was introduced by Neyman [1959], and discussed further, with specific reference to its relationship to two-step estimation, by Godfrey [1978] and Smith [1982]. One can see from the definition in equation <2.11.1.1> that the C-alpha statistic can be computed as n times the explained squared norm from the regression of

$$\mathbf{P}_n(\alpha^*)\mathbf{I}_n^{-1}(\theta^*)n^{-1}D_\theta \mathbf{l}_n(y;\theta^*)$$

on $\Phi(\beta^*)$ in the metric of $\mathbf{I}_n(\theta^*)$: this depends on the property that

$$\mathbf{I}_n^{-1}(\theta)\mathbf{P}_n'(\alpha) = \mathbf{P}_n(\alpha)\mathbf{I}_n^{-1}(\theta).$$

A more useful expression for this statistic stems from

$$\begin{aligned}\mathbf{P}_n\Phi(\Phi'\mathbf{I}_n\Phi)^{-1}\Phi'\mathbf{I}_n &= \Phi(\Phi'\mathbf{I}_n\Phi)^{-1}\Phi'\mathbf{I}_n - \Theta(\Theta'\mathbf{I}_n\Theta)^{-1}\Theta'\mathbf{I}_n\Phi(\Phi'\mathbf{I}_n\Phi)^{-1}\Phi'\mathbf{I}_n \\ &= \Phi(\Phi'\mathbf{I}_n\Phi)^{-1}\Phi'\mathbf{I}_n - \Theta(\Theta'\mathbf{I}_n\Theta)^{-1}\Theta'\mathbf{I}_n \\ &= \mathbf{P}_n - \mathbf{P}_{\Phi n},\end{aligned}$$

where the matrices $\mathbf{P}_n$, $\mathbf{P}_{\Phi n}$ are defined by equations <2.4.2.3> and <2.4.1.9> respectively. The truth of this result depends on the facts that at α^*,

$$\beta^* = \lambda(\alpha^*), \quad \theta^* = \phi(\beta^*),$$

$$D_\alpha\theta(\alpha^*) = \Phi(\beta^*)\Lambda(\alpha^*).$$

The expression

$$\mathbf{P}_n\Phi(\Phi'\mathbf{I}_n\Phi)^{-1}\Phi'\mathbf{P}_n'$$

in the centre of the C-alpha statistic <2.11.1.1> is then

$$\begin{aligned}(\mathbf{P}_n - \mathbf{P}_{\Phi n})\mathbf{I}_n^{-1}\mathbf{P}_n' &= (\mathbf{P}_n - \mathbf{P}_{\Phi n})\mathbf{P}_n\mathbf{I}_n^{-1}\\ &= (\mathbf{P}_n - \mathbf{P}_{\Phi n})\mathbf{I}_n^{-1}.\end{aligned}$$

Since

$$\begin{aligned}(\mathbf{P}_n - \mathbf{P}_{\Phi n})\mathbf{I}_n^{-1}(\mathbf{P}_n - \mathbf{P}_{\Phi n})' &= \mathbf{I}_n^{-1}(\mathbf{P}_n - \mathbf{P}_{\Phi n})'\mathbf{I}_n(\mathbf{P}_n - \mathbf{P}_{\Phi n})\mathbf{I}_n^{-1}\\ &= \mathbf{I}_n^{-1}(\mathbf{P}_n'\mathbf{I}_n\mathbf{P}_n - \mathbf{P}_{\Phi n}'\mathbf{I}_n\mathbf{P}_{\Phi n})\mathbf{I}_n^{-1},\end{aligned}$$

the statistic <2.11.1.1> can be written as

$$\begin{aligned}CA = n^{-1}D_\theta\mathbf{l}_n'(y;\theta^*)\mathbf{I}_n^{-1}(\theta^*)\,[&\mathbf{P}_n'(\alpha^*)\mathbf{I}_n(\theta^*)\mathbf{P}_n(\alpha^*)\\ &- \mathbf{P}_{\Phi n}'(\beta^*)\mathbf{I}_n(\theta^*)\mathbf{P}_{\Phi n}(\beta^*)]\,\mathbf{I}_n^{-1}D_\theta\mathbf{l}_n(y;\theta^*)\end{aligned} \qquad \text{<2.11.1.2>}$$

which can be seen to be the difference of the residual squared norms in the regressions of

$$\mathbf{I}_n^{-1}(\theta^*)n^{-1}D_\theta\mathbf{l}_n(y;\theta^*)$$

on $\Theta(\alpha^*)$ and on $\Phi(\alpha^*)$ respectively, in the metric of $\mathbf{I}_n(\theta^*)$. More directly, this can be expressed as

$$CA = n(\mathrm{RSN}(\Theta) - \mathrm{RSN}(\Phi)), \qquad \text{<2.11.1.3>}$$

in an obvious notation.

<u>2.11.2.</u> In the case where the alternative hypothesis model is just identified, that is,

$$D_\beta\phi(\beta) = \Phi(\beta)$$

may be taken to be square and non-singular, it is clear that the C-alpha statistic collapses to the expression

$$CA = n^{-1} D_\theta l_n'(y;\theta^*) \mathbf{P}_n'(\alpha^*) \mathbf{I}_n^{-1}(\theta^*) \mathbf{P}_n(\alpha^*) D_\theta l_n(y;\theta^*) \qquad \langle 2.11.2.1 \rangle$$

in just the same way that the Lagrange Multiplier statistic collapses in the same circumstances - see equation <2.9.3.1>. Note the presence still of the projection matrix $\mathbf{P}_n(\alpha^*)$: this seems to be the characteristic feature of the C-alpha test statistic.

The test statistic above can be interpreted as simply n times the residual squared norm of the regression of $n^{-1}\mathbf{I}_n^{-1}(\theta^*) D_\theta l_n(y;\theta^*)$ on $\Theta(\alpha^*)$ in the metric of $\mathbf{I}_n(\theta^*)$: that is, the first term in the expressions <2.11.1.2> or <2.11.1.3>.

It is interesting that the regression coefficient vector in this regression,

$$n^{-1}[\Theta'(\alpha^*)\mathbf{I}_n(\theta^*)\Theta(\alpha^*)]^{-1}\Theta'(\alpha^*)\mathbf{I}_n(\theta^*)\mathbf{I}_n^{-1}(\theta^*) D_\theta l_n(y;\theta^*),$$

is precisely the "update" term in the expression for the two-step estimator of α, $\hat{\alpha}$, given in equation <2.6.2.3>.

2.12. Difference Statistics.

2.12.1. Some of the test statistics for testing the hypotheses of equations <2.1.1.2> and <2.1.1.1>,

$\mathbf{H}_0$: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$

$\mathbf{H}_1$: $\theta = \phi(\beta)$,

(where the dimensions of θ and β, r_1 and s_0, differ) decompose naturally into the difference of a test statistic for a test of the hypotheses

$$\mathbf{H}_0: \quad \theta = \phi[\lambda(\alpha)] = \theta(\alpha) \qquad \text{<2.12.1.1>}$$

$$\mathbf{H}_0^*: \quad \theta \neq \theta(\alpha) \qquad \text{<2.12.1.2>}$$

and a test statistic for the hypotheses

$$\mathbf{H}_1: \quad \theta = \phi(\beta) \qquad \text{<2.12.1.3>}$$

$$\mathbf{H}_1^*: \quad \theta \neq \phi(\beta): \qquad \text{<2.12.1.4>}$$

that is, where $\mathbf{H}_0^*$ and $\mathbf{H}_1^*$ are equivalent to saying that θ is unrestricted. Thus, the maximum likelihood estimator of θ under both $\mathbf{H}_0^*$ and $\mathbf{H}_1^*$ is $\hat{\theta}$, whilst under $\mathbf{H}_1$ it is $\tilde{\phi}$, and under $\mathbf{H}_0$ it is $\tilde{\theta}$.

The Likelihood Ratio statistic of section 2.8. clearly decomposes in this way:

$$LR = -2[l_n(y;\tilde{\theta}) - l_n(y;\tilde{\phi})]$$
$$= -2\{[l_n(y;\tilde{\theta})-l_n(y;\hat{\theta})] - [l_n(y;\tilde{\phi})-l_n(y;\hat{\theta})]\};$$

in fact, this decomposition was the basis for establishing the well known limit distribution (under the null hypothesis) $LR \overset{a}{\sim} \chi^2_{r_1-r_0}$.

2.12.2 It is useful to consider whether the Lagrange

Multiplier statistic for testing the null hypothesis $\mathbf{H}_0$ of equation <2.12.1.1> against the alternative hypothesis $\mathbf{H}_1$ of equation <2.12.1.3> can be decomposed in this way; the test statistic is given by equation <2.9.2.1>,

$$LM = n^{-1}D_\theta \mathbf{l}_n'(y;\tilde{\theta})\Phi(\tilde{\beta}_0)\,[\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi(\tilde{\beta}_0)]^{-1}\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})D_\theta \mathbf{l}_n(y;\tilde{\theta}),$$

and from subsection 2.9.1.,

$$LM \overset{a}{\approx} n^{-1}D_\theta \mathbf{l}_n'(y;\theta^0)\mathbf{P}(\alpha^0)\Phi(\beta^0)\,[\Phi'(\beta^0)\mathbf{I}(\theta^0)\Phi(\beta^0)]^{-1}$$
$$\times\ \Phi'(\beta^0)\mathbf{P}'(\alpha^0)D_\theta \mathbf{l}_n(y;\theta^0)$$
$$\overset{a}{\approx} n^{-1}D_\theta \mathbf{l}_n'(y;\theta^0)(\mathbf{P}(\alpha^0) - \mathbf{P}_\Phi(\beta^0))\mathbf{I}^{-1}(\theta^0)D_\theta \mathbf{l}_n(y;\theta^0).$$

When the alternative hypothesis is unrestricted, as in subsection 2.9.3., the Lagrange Multiplier statistic is, say,

$$LM_0 = n^{-1}D_\theta \mathbf{l}_n'(y;\tilde{\theta})\mathbf{I}_n^{-1}(\tilde{\theta})D_\theta \mathbf{l}_n(y;\tilde{\theta})$$
$$\overset{a}{\approx} n^{-1}D_\theta \mathbf{l}_n'(y;\theta^0)\mathbf{P}(\alpha^0)\mathbf{I}^{-1}(\theta^0)\mathbf{P}'(\alpha^0)D_\theta \mathbf{l}_n(y;\theta^0),$$

which is therefore appropriate for a test of the null hypothesis $\mathbf{H}_0$ of equation <2.12.1.1> against the alternative hypothesis $\mathbf{H}_0^*$ of equation <2.12.1.2>. One can directly convert this latter statistic, by a notation switch, into the Lagrange Multiplier statistic appropriate for a test of the "null" hypothesis $\mathbf{H}_1$ of equation <2.12.1.3> against the alternative $\mathbf{H}_1^*$ of equation <2.12.1.4>, since $\tilde{\phi}$, $\tilde{\beta}$ are the maximum likelihood estimators under the hypothesis $\mathbf{H}_1$:

$$LM_1 = n^{-1}D_\theta \mathbf{l}_n'(y;\tilde{\phi})\mathbf{I}_n^{-1}(\tilde{\phi})D_\theta \mathbf{l}_n(y;\tilde{\phi}),$$

and, by the same arguments which verify the limiting distribution of LM_0,

$$LM_1 \overset{a}{\approx} n^{-1}D_\theta \mathbf{l}_n'(y;\theta^0)\mathbf{P}_\Phi(\beta^0)\mathbf{I}^{-1}(\theta^0)\mathbf{P}_\Phi'(\beta^0)D_\theta \mathbf{l}_n(y;\theta^0),$$

where $\mathbf{P}_\Phi(\beta)$ is defined by equation <2.4.1.9>. Recalling that

$$\mathbf{P}\mathbf{I}^{-1} = \mathbf{I}^{-1}\mathbf{P}', \qquad \mathbf{P}_\Phi\mathbf{I}^{-1} = \mathbf{I}^{-1}\mathbf{P}_\Phi',$$

it then follows (deleting the dependence of $\mathbf{P}$, $\mathbf{P}_{\Phi}$ and $\mathbf{I}$ on α, β, and θ) that

$$\begin{aligned} LM_0 - LM_1 &\overset{a}{\approx} n^{-1}D_\theta l_n'(y;\theta^0)[\mathbf{P}\mathbf{I}^{-1}\mathbf{P}' - \mathbf{P}_{\Phi}\mathbf{I}^{-1}\mathbf{P}_{\Phi}']D_\theta l_n(y;\theta^0) \\ &= n^{-1}D_\theta l_n'(y;\theta^0)(\mathbf{P} - \mathbf{P}_{\Phi})\mathbf{I}^{-1}D_\theta l_n(y;\theta^0) \\ &\overset{a}{\approx} LM. \end{aligned}$$

Since LM, LM_0, and LM_1 depend on different estimators, there is no reason why equality should hold in

$$LM = LM_0 - LM_1,$$

but there is an "asymptotic" equivalence.

2.12.3. With this result in mind, and the results showing the asymptotic equivalence of Wald and C-alpha statistics with the LM statistic under the null hypothesis $\mathbf{H}_0$ of equation <2.12.1.1>, one can examine the possibility of the decomposition of a Wald test statistic for testing <2.12.1.1> against <2.12.1.3> into the difference of Wald test statistics for tests of <2.12.1.1> against <2.12.1.2>, and <2.12.1.3> against <2.12.1.4>.

From subsection 2.10.2., equation <2.10.2.1>, the Wald statistic, now denoted W_0,

$$W_0 = n(\hat{\theta} - \theta(\alpha^*))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha^*)),$$

α^* being the minimum chi-squared estimator of α, is appropriate for testing the null hypothesis $\mathbf{H}_0$ of equation <2.12.1.1> against the alternative hypothesis $\mathbf{H}_0^*$ of equation <2.12.1.2>; one can also define an analogous minimum chi-squared estimator of β, say, β^*, under the hypothesis $\mathbf{H}_1$

of equation <2.12.1.3>, and which will result in a Wald statistic,

$$W_1 = n(\hat{\theta} - \phi(\beta^*))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \phi(\beta^*)).$$

The analysis of subsection 2.10.2. shows that these two statistics are asymptotically equivalent to LM_0 and LM_1 respectively, as defined in the previous subsection, so that

$$LM \overset{a}{\approx} LM_0 - LM_1 \overset{a}{\approx} W_0 - W_1;$$

from subsection 2.10.3, the Wald statistic for testing <2.12.1.1> against <2.12.1.3> is

$$W = n(\tilde{\beta} - \lambda(\alpha^*))'\Phi'(\tilde{\beta})\mathbf{I}_n(\tilde{\phi})\Phi(\tilde{\beta})(\tilde{\beta} - \lambda(\alpha^*))$$

and

$$W \overset{a}{\approx} LM.$$

Hence,

$$W \overset{a}{\approx} W_0 - W_1,$$

although as in the case of the Lagrange Multiplier test statistics,

$$W \neq W_0 - W_1.$$

The same type of asymptotic equivalence can also be established for the Wald test statistics based on the linearised minimum chi-squared estimators of subsection 2.10.4.; even when the same initial estimators are used, a finite sample equivalence cannot be obtained. These results will not be shown formally.

<u>2.12.4.</u> In the case of the C-alpha test statistic for testing the null hypothesis $\mathbf{H}_0$ of equation <2.12.1.1> against

the alternative hypothesis $\mathbf{H}_1$ of equation <2.12.1.3>, the expression of equation <2.11.1.2> can be adapted slightly to

$$CA = n^{-1}D_\theta l_n'(y;\theta^*)\mathbf{I}_n^{-1}(\theta^*)[\mathbf{P}_n(\alpha^*) - \mathbf{P}_{\Phi n}(\beta^*)]'\mathbf{I}_n^{-1}(\theta^*)D_\theta l_n(y;\theta^*).$$

The expression given in equation <2.11.2.1>, here denoted CA_0,

$$CA_0 = n^{-1}D_\theta' l_n(y;\theta^*)\mathbf{I}_n^{-1}(\theta^*)\mathbf{P}_n'(\alpha^*)\mathbf{I}_n^{-1}(\theta^*)D_\theta l_n(y;\theta^*)$$

can be seen as a suitable test statistic for a test of the null hypothesis $\mathbf{H}_0$ of equation <2.12.1.1> against the alternative $\mathbf{H}_0^*$ of equation <2.12.1.2>, whilst there is a natural definition of a statistic CA_1 for a test of <2.12.1.3> against <2.12.1.4>:

$$CA_1 = n^{-1}D_\theta l_n'(y;\theta^*)\mathbf{I}_n^{-1}(\theta^*)\mathbf{P}_{\Phi n}'(\beta^*)\mathbf{I}_n^{-1}(\theta^*)D_\theta l_n(y;\theta^*),$$

and this pair of definitions shows that

$$CA \equiv CA_0 - CA_1.$$

Notice that this equality only holds when the same initial estimators

$$\alpha^*,\ \beta^* = \lambda(\alpha^*),\ \theta^* = \phi(\beta^*)$$

are used to obtain the two-step estimators under the hypotheses $\mathbf{H}_0$ and $\mathbf{H}_1$. As soon as this condition fails, the equivalence above is only asymptotic.

<u>2.12.5</u> The use of these "difference" statistics was discussed by Aitchison [1962], in the context of a constraint equation formulation; he argued that it might be convenient in some cases to use "hybrid" difference statistics. For example, one might use

$$LR_0 - W_1,$$

where LR_0 is the Likelihood Ratio statistic for a test of the null hypothesis <2.12.1.1> against the alternative $\mathbf{H}_0^*$ of equation <2.12.1.2>, and W_1 the Wald statistic for a test of the hypothesis $\mathbf{H}_1$ of equation <2.12.1.3> against the alternative $\mathbf{H}_1^*$ of equation <2.12.1.4>. However, such cases will not be discussed in what follows, in order to keep some bounds on the number of different test statistics that can be created.

2.13. An Overview and Evaluation.

In this section, the similarities and differences between the various test statistics are discussed, along with some observations on the circumstances under which a particular test statistic may be preferred. Finally, the test statistics are classified according to whether they are tests of specification or misspecification.

2.13.1. Of the various test statistics, only the Likelihood Ratio and the Wald statistics are the values (or differences) of a criterion function whose optimisation yields the corresponding estimators. This is useful if one desires to have an estimation procedure that will yield the required test statistics "free of charge"; the C-alpha statistic of equation <2.11.2.1>, for the case where the alternative hypothesis <2.1.1.1> is just identified, also has this property.

This observation leads on to the similarities between Wald and C-alpha test statistics: strictly speaking, Wald tests based on linearised minimum chi-squared estimators. This similarity may be seen most easily by comparing equations <2.6.2.3> and <2.7.1.2>. The former gives the two-step estimator of α,

$$\hat{\alpha} = \alpha^* + (\theta'(\alpha^*)I_n(\theta^*)\theta(\alpha^*))^{-1}\theta'(\alpha^*)n^{-1}D_\theta l_n(y;\theta^*)$$

whilst the latter equation gives the linearised minimum chi-squared estimator,

$\alpha^{\nabla} = \alpha^* + (\mathbf{B}'(\alpha^*)\mathbf{I}_n(\hat{\theta})\mathbf{B}(\alpha^*))^{-1}\mathbf{B}'(\alpha^*)\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta^*)$,

α^* and θ^* being initial estimators such that

$\theta^* = \theta(\alpha^*)$.

Apart from the use of $\mathbf{I}_n(\hat{\theta})$ in α^{∇}, it can be seen that

$\mathbf{I}_n^{-1}(\theta^*)n^{-1}D_\theta l_n(y;\theta^*)$ and $(\hat{\theta} - \theta^*)$

are approximations to each other. Where a C-alpha statistic uses a regression with regressand

$\mathbf{I}_n^{-1}(\theta^*)n^{-1}D_\theta l_n(y;\theta^*)$

and metric $\mathbf{I}_n(\theta^*)$, the corresponding linearised minimum chi-square Wald test statistic uses $(\hat{\theta} - \theta^*)$ as the regressand, with metric $\mathbf{I}_n(\hat{\theta})$.

Another aspect of this discussion is that if the alternative hypothesis is just identified, the C-alpha test procedure would fail if the unrestricted maximum likelihood estimator $\hat{\theta}$ were taken as θ^*, with α^* any solution to

$\hat{\theta} = \theta(\alpha)$.

In the Wald test procedure, the use of the efficient unrestricted estimator $\hat{\theta}$ is essential. One can, in fact, base a Wald test on an inefficient estimator (relative to maximum likelihood) of α by using an inefficient estimator of θ, and using as the metric of minimisation, the inverse of an estimate of its limiting covariance matrix. The Wald test statistic so obtained will not, however, be asymptotically equivalent to the Likelihood Ratio statistic. Such an estimator, with the associated test statistic has had application in regression models with truncated normal errors.

A Wald test statistic is usually seen as being dependent only on estimation under the alternative hypothesis. The failure of the proposed Wald test statistics to satisfy this property really arises from the combination of the constraint parameter formulation and the linearisation involved; this criticism does not apply to the Wald statistics based on the full minimum chi-squared estimation principle. The C-alpha statistics are effectively versions of the Lagrange Multiplier statistic, and thus only require estimation under the null hypothesis.

In any case, when the alternative hypothesis <2.1.1.1> is overidentified, the use of "difference" statistics, discussed in section 2.12., will at the very least require estimation under both the null and alternative hypotheses, whatever test statistic is used.

<u>2.13.2.</u> This leads naturally on to the question of whether there are any circumstances in which a particular test statistic may be preferred. There are a number of factors to consider, but perhaps the most basic one is the nature of the computing facilities available to the researcher. A maximum likelihood estimation program will produce "free" test statistics, which is quite useful provided that the "job turnround" is sufficiently fast for "learning from the data" to proceed satisfactorily. If turnround is slow, and/or computation costly, a simpler estimator like the minimum chi-squared or two-step may be better, and will still produce

"free" test statistics.

Again, if the more restricted model is easy to estimate, a Lagrange Multiplier or C-alpha test may be appropriate: the test statistic simply requires a regression using quantities evaluated during the estimation.

This leads into another consideration: is estimation under both the null and alternative hypotheses of interest? To this can be added, where appropriate, the need to compute an unrestricted estimator of θ. The answer to such questions will clearly vary tremendously with circumstances: the discussion in the previous subsection shows that if estimation under only one of the models is strictly required, a suitable test statistic is readily available - Wald or Lagrange Multiplier, depending on whether estimation under the null or alternative hypotheses is required. If estimation under both models is required, virtually all of the test statistics discussed in this chapter are available.

Another situation, not discussed fully here, is where there is a nested sequence of hypotheses to be tested: Lagrange Multiplier statistics are less attractive here, since estimators from one stage cannot be employed directly at the ensuing stage. The minimum chi-squared principle and the associated Wald test procedure might be useful if there is an easily obtained unrestricted estimator under the overall null hypothesis. Likelihood Ratio statistics are another natural possibility in this context.

2.13.3. In the Introduction to Chapter 1 (section 1.1.) there was considerable discussion of the nature of tests of "specification" and "misspecification": the former was seen as "confirming one's theory", whilst the latter was described as "detecting errors in one's model". It was, however, also noted that the imposition of incorrect restrictions could be seen as a misspecification, which tends to obscure the distinction. Another aspect of the detection of misspecification is that if the null hypothesis is rejected, it may not be clear what the specific alternative hypothesis model is: a specification test usually takes the form of a test of further restrictions to a well-specified model.

In addition, a further distinction between "symmetric" and "asymmetric" tests was made, in the case where the alternative hypothesis <2.1.1.1> is just identified: a symmetric test statistic is one that simply needs the information that θ is unrestricted under this hypothesis, whilst an asymmetric test is one that needs the information that

$\theta = \phi(\beta)$

under this hypothesis. One can see that a symmetric test is also a misspecification test, and may also be a specification test, dependent on the information available to the investigator, whilst an asymmetric test is clearly a specification test.

All of the test statistics discussed in this chapter are symmetric in the sense just used; some of them (the Lagrange Multiplier and C-alpha) do not even need the unrestricted estimator $\hat{\theta}$, unlike the Likelihood Ratio and Wald tests. When the alternative hypothesis is overidentified, all of the test statistics are asymmetric, and hence are tests of specification.

One is lead inevitably to the conclusion that the misspecification-specification test distinction is useful, but not very useful, in the kind of statistical models discussed in this Chapter: too much information, namely, the knowledge of the log-likelihood function of the observable random vectors $y_1,\ldots y_n$, has been assumed to be available. Use of an "incorrect" log-likelihood function is a misspecification that one would be powerless to detect by the methods of this Chapter; however, the use of test statistics for non-nested hypotheses, discussed in Chapters 7-9, may resolve this difficulty.

Chapter 3 : Estimation and Identification in the Linear Simultaneous Equations Model.

3.1. Introduction and Outline.

3.1.1. In this Chapter, the estimation and identification of two linear simultaneous equations models is considered, under the supposition that one of them is the "true" model, generating the observed random vectors $y_1,\ldots,y_n$. The basic aim of this Chapter is to prepare for the discussion of tests of overidentifying restrictions in Chapter 5, following on from the general results on maximum likelihood, two-step and minimum chi-squared estimation in the previous Chapter. In the penultimate section of this Chapter, limited information maximum likelihood estimation is discussed, and this leads naturally into the tests of identification discussed in the next Chapter.

Since the chief aim of this work is inference in the simultaneous equations model, it will be necessary to define the simultaneous equations models supposed to rule under each competing hypothesis: it is convenient to repeat here the definitions given in subsection 1.2.2. (for the alternative hypothesis model) and subsection 1.3.3. (for the null hypothesis model). It is also useful to recall the convention that dimensions, error terms, parameters and some parameter estimators attached to the alternative hypothesis model are subscripted "1", whilst a subscript "0" is used for the same purpose in the null hypothesis model.

<u>3.1.2.</u> Under the alternative hypothesis, the $m \times 1$ independent random vectors $y_1,\dots,y_n$ are assumed to be generated by the reduced form

$$y_t = \Pi_1' x_t + v_{1t}, \quad t = 1,\dots,n.$$

Here, v_{1t} is an $m \times 1$ random vector with mean vector 0 and covariance matrix Ω_1, x_t a $k_1 \times 1$ vector of nonstochastic, linearly independent and bounded exogenous variables, whilst Π_1 is a $k_1 \times m$ matrix of parameters, generated via the "structural form"

$$A_1' y_t + B_1' x_t = u_{1t}, \quad t = 1,\dots,n$$

as

$$\Pi_1 = - B_1 A_1^{-1},$$

so that there is an implicit assumption that the $m \times m$ matrix A_1 is nonsingular. The structural form disturbance vector u_{1t} is defined by

$$u_{1t} = A_1' v_{1t}, \quad t = 1,\dots,n,$$

and hence has mean 0 and covariance matrix

$$\Sigma_1 = A_1' \Omega_1 A_1. \qquad \text{<3.1.2.1>}$$

Observation matrices Y, X, U_1 and V_1 can be defined as follows:

$$Y' = (y_1,\dots,y_n); \quad X' = (x_1,\dots,x_n); \quad U_1' = (u_{11},\dots,u_{1n});$$

$$V_1' = (v_{11},\dots,v_{1n}),$$

X being $n \times k_1$, whilst Y, V_1, U_1 are $n \times m$. The reduced form can then be expressed as

$$Y = X\Pi_1 + V_1,$$

and the structural form as

$$YA_1 + XB_1 = U_1. \tag{3.1.2.2}$$

It is convenient to define the $n \times (m + k_1)$ matrix Z_1 as

$$Z_1 = (Y, X) \tag{3.1.2.3}$$

and the $(m + k_1) \times m$ matrix C_1 as

$$C_1' = (A_1', B_1'). \tag{3.1.2.4}$$

One can then represent the structural form model <3.1.2.2> as

$$Z_1C_1 = U_1,$$

or in "long vector" form as

$$(I_m \otimes Z_1)\text{vec } C_1 = \text{vec } U_1.$$

Let

$$g_1 = \text{vec } C_1, \quad u_1 = \text{vec } U_1: \tag{3.1.2.5}$$

then, <3.1.2.2> is equivalent to

$$(I_m \otimes Z_1)g_1 = u_1; \tag{3.1.2.6}$$

the covariance matrix of the vector u_1 is

$$\Sigma_1 \otimes I_n.$$

Under this hypothesis, the structural form parameters contained in the vector g_1 are supposed to satisfy the restrictions

$$g_1 = K\delta + k, \tag{3.1.2.7}$$

where K is a known $m(m + k_1) \times q_1$ matrix with full column rank, and k a known $m(m + k_1) \times 1$ vector; δ is the vector of unrestricted or "free" structural parameters.

<u>3.1.3.</u> The null hypothesis model can be approached from two points of view: the first is simply as a more restricted version of the alternative hypothesis model defined by <3.1.2.6> and <3.1.2.7>, in which the parameter vector δ

satisfies

$$\delta = L\gamma + r, \qquad \text{<3.1.3.1>}$$

where L is a known $q_1 \times q_0$ matrix with full column rank, r a known $q_1 \times 1$ vector, and γ a $q_0 \times 1$ vector of free parameters. The other approach is to write out the null hypothesis model in full as a simultaneous equations model, with reduced form

$$y_t = \Pi_0' x_t + v_{0t}, \quad t = 1,\ldots,n,$$

or

$$Y = X\Pi_0 + V_0, \qquad \text{<3.1.3.2>}$$

where

$$V_0' = (v_{01},\ldots,v_{0n}),$$

and the structural form is

$$A_0' y_t + B_0' x_t = u_{0t}, \quad t = 1,\ldots,n,$$

or

$$YA_0 + XB_0 = U_0. \qquad \text{<3.1.3.3>}$$

Here, the random vector v_{0t} has mean 0 and covariance matrix Ω_0, so that

$$u_{0t} = A_0' v_{0t}$$

has mean 0 and covariance matrix

$$\Sigma_0 = A_0' \Omega_0 A_0;$$

A_0 is $m \times m$ and non-singular, so that

$$\Pi_0 = - B_0 A_0^{-1}.$$

Defining

$$C_0' = (A_0', B_0'),$$

and

$$g_0 = \text{vec } C_0, \qquad \text{<3.1.3.4>}$$

the structural parameters are supposed to satisfy

$g_0 = H\gamma + h,$

where H is a known $m(m + k_1) \times q_0$ matrix of full column rank, and h a known $m(m + k_1) \times 1$ vector. One could thus write, analogously to equations <3.1.2.6> and <3.1.2.7>,

$$(I_m \otimes Z_1)g_0 = u_0, \qquad \text{<3.1.3.5>}$$

$$g_0 = H\gamma + h, \qquad \text{<3.1.3.6>}$$

where

$u_0 = \text{vec } U_0$

has covariance matrix $\Sigma_0 \otimes I_n$.

<u>3.1.4.</u> Viewing this null hypothesis model as a more restricted version of the alternative hypothesis model, the unrestricted structural form parameters g_1 satisfy

$g_1 = K\delta + k,$

$\delta = L\gamma + r,$

or,

$$g_1 = K(L\gamma + r) + k$$
$$= KL\gamma + (Kr + k)$$

so that H and h are defined by

$$H = KL, \qquad \text{<3.1.4.1>}$$

$$h = Kr + k, \qquad \text{<3.1.4.2>}$$

in the manner described in subsection 1.3.2. . The switch in notation here from g_1 to g_0 in equation <3.1.3.6> is possibly confusing, but merely arises from the desire to label parameters and hence estimators as coming from a particular model, even if different symbols represent the same parameter. Such a distinction will turn out to be very useful in the discussion of simultaneous equations models that are "non-nested" with respect to each other, and preserves notational consistency throughout the thesis.

To emphasise the point, let γ^0 be the true value of γ ruling in the null hypothesis model: the corresponding true values of g_0 and δ are denoted for the moment by g_0^0 and δ^0 respectively, and are given by

$$g_0^0 = H\gamma^0 + h, \qquad \delta^0 = L\gamma^0 + r,$$

so that the true value of the vector g_1 under the null hypothesis, denoted g_1^0, is

$$\begin{aligned} g_1^0 &= K(L\gamma^0 + r) + k \\ &= KL\gamma^0 + (Kr + k) \\ &= g_0^0. \end{aligned}$$

Given this, the true value of g_0 or g_1 under the null hypothesis will be denoted g^0, and similarly for the other parameters:

$$\Pi^0,\ A^0,\ B^0,\ C^0,\ \Omega^0,\ \Sigma^0$$

are the true values of

$$A_0 \text{ and } A_1, \quad B_0 \text{ and } B_1, \quad C_0 \text{ and } C_1, \quad \Omega_0 \text{ and } \Omega_1, \quad \Sigma_0 \text{ and } \Sigma_1$$

respectively, under the null hypothesis. This convention will also be applied to the $k_1 \times m(m + k_1)$ matrices

$$\mathbf{D}_0 = (\Pi_0, I_{k_1}), \qquad \mathbf{D}_1 = (\Pi_1, I_{k_1}),$$

whose true values under the null hypothesis equal

$$\mathbf{D}^0 = (\Pi^0, I_{k_1}).$$

3.2. Identification.

3.2.1. In the previous Chapter, (section 2.2.), the strong consistency of maximum likelihood estimators of the parameter vector θ in the scaled log-likelihood function

$$n^{-1}l_n(y;\theta)$$

depended on the assumption that the uniform almost sure limit of this function had a unique maximum at the true value θ^0; when θ depends on a vector α,

$$\theta = \theta(\alpha),$$

(as in equation <2.1.1.3>), strong consistency for the maximum likelihood estimator of α depends on the existence of a unique solution for α in

$$\theta^0 = \theta(\alpha):$$

if so, α is said to be uniquely identified (at θ^0).

This situation was described as the "null hypothesis" model in Chapter 2, and corresponding assumptions led to to the strong consistency (given the truth of the null hypothesis) of the maximum likelihood estimators of the parameters θ and β in the alternative hypothesis model of equation <2.1.1.1>, where

$$\theta = \phi(\beta).$$

In embedding the linear simultaneous equations model in this general constrained maximum likelihood problem, as in subsection 1.3.3., the parameter vector θ is taken to be

$$\theta = \begin{bmatrix} v(\Omega_0) \\ \text{vec } \Pi_0 \end{bmatrix},$$

whilst

$$\alpha = \begin{bmatrix} v(\Omega_0) \\ \gamma \end{bmatrix},$$

and similarly,

$$\phi = \begin{bmatrix} v(\Omega_1) \\ \text{vec } \Pi_1 \end{bmatrix}, \qquad \beta = \begin{bmatrix} v(\Omega_1) \\ \delta \end{bmatrix},$$

Comparing θ and α, it is clear that demanding a unique solution for α for a given θ is equivalent to demanding that a unique γ-value corresponds to a given Π_0-value. Similarly, a unique value for δ has to correspond to a given value for Π_1.

3.2.2. This discussion permits a link with the traditional econometric discussion of "observational equivalence" and identification: it will be convenient to discuss this in the context of the alternative hypothesis model defined by equations <3.1.2.2>,

$$YA_1 + XB_1 = U_1$$

and <3.1.2.7>,

$$g_1 = \text{vec}\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = K\delta + k.$$

Define a "structure" S_1 to be a value of the triple $\{A_1, B_1, \Sigma_1\}$: because of the dependence of A_1, B_1, Σ_1 on $v(\Omega_1)$ and δ, that is, the vector β, the set of structures M_1 corresponds to the set of all possible values of the vector β.

Let S_1^*, $S_1^{\#}$ be possible structures: they are said to be observationally equivalent if they imply the same value of the vector θ: that is, the same mean vector (for given $x_1,\ldots,x_n$ values) and covariance matrix for the observable random vectors $y_1,\ldots,y_n$ in the simultaneous equations model. This is a "wide-sense" definition of observational equivalence; more generally, one would say that two structures are observationally equivalent if they yield the same joint probability distribution for $y_1,\ldots,y_n$. It is well known that in the simultaneous equations model, two structures

$$S_1^{\#} = \{A_1^{\#},\ B_1^{\#},\ \Sigma_1^{\#}\}$$

$$S_1^* = \{A_1^*,\ B_1^*,\ \Sigma_1^*\}$$

are (wide sense) observationally equivalent if and only if there exists a nonsingular matrix $\mathbf{R}$ such that

$$A_1^{\#} = A_1^*\mathbf{R}, \quad B_1^{\#} = B_1^*\mathbf{R}, \quad \Sigma_1^{\#} = \mathbf{R}'\Sigma_1^*\mathbf{R}.$$

One can then go on to say that a structure $S_1^{\#} \in M_1$ is identified if and only if there exists no other structure $S_1^* \in M_1$ which is observationally equivalent to $S_1^{\#}$: it is clear that if M_1 were defined as

$$M_1 = \left\{S_1 \middle| S_1 = \{A_1,\ B_1,\ \Sigma_1\},\ A_1 \text{ nonsingular},\ \Sigma_1 \text{ positive definite}\right\},$$

no $S_1 \in M_1$ would be identified. This can be seen more easily from the relationship

$$\Pi_1 A_1 + B_1 = 0,$$

or, defining

$$\Omega_1 = (\Pi_1,\ I_{k_1}),$$

this is equivalent to

$\Omega_1 C_1 = 0$,

or to

$$(I_m \otimes \Omega_1) g_1 = 0: \qquad \langle 3.2.2.1\rangle$$

that is,

$g_1 \in N(I_m \otimes \Omega_1)$.

The demand in equation <3.1.2.7> that

$g_1 = K\delta + k$

may restrict this null space to be empty, to contain exactly one vector, or to contain several linearly independent vectors. Combining equation <3.1.2.7> and <3.2.2.1>, one obtains

$(I_m \otimes \Omega_1)K\delta = -(I_m \otimes \Omega_1)k$:

δ is identified if and only if this equation has a unique solution, given a value for Ω_1. The conditions for this are that the equations are consistent, and that

rank $(I_m \otimes \Omega_1)K = q_1 \leq mk_1$

(since K has q_1 columns). This type of rank condition has been given by Monfort [1978] and implicitly by Rothenberg [1971].

Under the alternative hypothesis model, it is supposed that the set of structures satisfying <3.1.2.7> is not empty, and hence the consistency demand is redundant if the rank condition is satisfied, as it will be in any case if

$q_1 = mk_1$,

when the dimensionality of β matches that of θ, since Π_1 is $k_1 \times m$. In this case, the model is described as just-identified, whereas when $q_1 < mk_1$, it is usually described as over-identified.

<u>3.2.3.</u> The null hypothesis model can be obtained from the alternative hypothesis model by imposing the additional restrictions

$$\delta = L\gamma + r:$$

see equation <3.1.3.1>. Suppose that δ is identified given a value of Π_1: then, γ is identified provided that L has full column rank q_0, which was assumed in subsection 3.1.3. .
Using the second method of defining the null hypothesis model, one can show, using the arguments of the preceding subsection, that γ is identified if and only if there is a unique solution to

$$(I_m \otimes \Omega_0)H\gamma = -(I_m \otimes \Omega_0)h,$$

that is, if and only if a solution exists and

$$\text{rank}(I_m \otimes \Omega_0)H = q_0 \leq mk_1.$$

3.3. Estimation in the Linear Simultaneous Equations Model

3.3.1. This section obtains the maximum likelihood estimators of the parameters

$v(\Omega_1)$, vec Π_1, δ

and

$v(\Omega_0)$, vec Π_0, γ

of the alternative and null hypothesis models defined in section 3.1., under a normality assumption for the reduced form error vectors v_{1t}, v_{0t}. More precisely,

$v_{1t} \sim N(0,\Omega_1)$, $\quad t = 1,\ldots,n,$

$v_{0t} \sim N(0,\Omega_0)$, $\quad t = 1,\ldots,n,$

whilst the independence assumption for these vectors is maintained.

There is a certain ambivalence in this thesis, as elsewhere in the econometric literature, about the role and effect of this normality assumption: the nature of the likelihood function for the observations $y_1,\ldots,y_n$ is thereby determined, but one may also use the corresponding "log-likelihood function" as a criterion function to be optimised in order to obtain "quasi-maximum likelihood" estimators. Such estimators share, under suitable assumptions, many of the large sample properties of the "full" maximum likelihood estimators. However, one curious feature of derivations of the limiting normal distributions for such quasi-maximum likelihood estimators is that the information matrix associated with normality is used. That is, the third and fourth moment properties of the multivariate normal distribution are employed; rarely does one see an explicit assumption that the error distribution is non-normal, but none the less shares these moment properties of the multivariate normal.

3.3.2 It is again convenient to start with estimaton of the parameters of the alternative hypothesis model

$$y_t = \Pi_1' x_t + v_{1t}, \quad t = 1,\ldots,n,$$

or

$$Y = X\Pi_1 + V_1,$$

where

$$\Pi_1 = -B_1 A_1^{-1}$$

and

$$g_1 = \mathrm{vec}\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \mathrm{vec}\, C_1 = K\delta + k.$$

With the normality assumption, the log-likelihood function for $y_1,\ldots,y_n$ is

$$\begin{aligned} l_n(y;\phi) &= nms - \tfrac{1}{2}n\log\det\Omega_1 - \tfrac{1}{2}\textstyle\sum_{t=1}^{n} v_{1t}'\Omega_1^{-1}v_{1t} \\ &= nms - \tfrac{1}{2}n\log\det\Omega_1 - \tfrac{1}{2}\mathrm{tr}(\Omega_1^{-1}(Y - X\Pi_1)'(Y - X\Pi_1)), \end{aligned} \qquad \langle 3.3.2.1\rangle$$

where s is defined in subsection 1.5.1. . To find the maximum likelihood estimators of Ω_1, Π_1 and δ, this has to maximised subject to the constraint

$$\phi = \begin{bmatrix} v(\Omega_1) \\ \mathrm{vec}\,\Pi_1 \end{bmatrix} = \phi\left\{\begin{bmatrix} v(\Omega_1) \\ \delta \end{bmatrix}\right\}.$$

It is clear that vec Π_1 depends only on δ, and trivially,

$v(\Omega_1)$ in ϕ depends only on $v(\Omega_1)$ in β; so, the function $\phi(\beta)$ splits into two subvector functions,

$$\phi(\beta) = \begin{bmatrix} \phi_1(v(\Omega_1)) \\ \phi_2(\delta) \end{bmatrix},$$

but where $\phi_1(.)$ is the identity function. Thus, the constraint

$$\phi = \phi(\beta)$$

amounts to

$$v(\Omega_1) = v(\Omega_1),$$

$$\text{vec } \Pi_1 = \phi_2(\delta);$$

since the first of these equations is satisfied trivially, the log-likelihood function <3.3.2.1> will be maximised only subject to

$$\text{vec } \Pi_1 = \phi_2(\delta)$$

for the purposes of finding the maximum likelihood estimators, although from the point of view of applying the formal limit distribution results of section 2.4., the zero Lagrange multiplier attached to the equation

$$v(\Omega_1) = v(\Omega_1)$$

will have to be included.

The Lagrange multiplier of the general problem, μ, may be partitioned to suit the partitioning of $\phi(.)$ into $\phi_1(.)$ and $\phi_2(.)$ as

$$\mu' = (\mu_1', \mu_2'),$$

and the argument above shows that

$$\mu_1 \equiv 0.$$

The dimensions of these Lagrange multipliers are $\frac{1}{2}m(m + 1)$ and mk_1 respectively.

So, to find the maximum likelihood estimators $\tilde{\Pi}_1$, $\tilde{\Omega}_1$, $\tilde{\delta}$, the stationary points of the Lagrangean

$$\mathcal{L}_1 = n^{-1}l_n(y;\phi) + \mu_2'(\text{vec } \Pi_1 - \phi_2(\delta)) \qquad \text{<3.3.2.2>}$$

will have to be found, and this will require the differentials of $n^{-1}l_n(y;\phi)$ and $\phi_2(\delta)$.

<u>3.3.3.</u> With $n^{-1}l_n(y;\phi)$ defined by equation <3.3.2.1>, its differential is

$$n^{-1}dl_n(y;\phi) = -\tfrac{1}{2}d\log\det\Omega_1 - \tfrac{1}{2}n^{-1}\text{tr}[d\Omega_1^{-1}(Y - X\Pi_1)'(Y - X\Pi_1)]$$
$$- \tfrac{1}{2}n^{-1}\text{tr}[\Omega_1^{-1}d\{(Y - X\Pi_1)'(Y - X\Pi_1)\}].$$

To find the various differentials in this expression, the results given in subsections 1.6.1. and 1.6.2. will be freely used: thus,

$$\begin{aligned} n^{-1}dl_n(y;\phi) &= -\tfrac{1}{2}\text{tr}[\Omega_1^{-1}d\Omega_1] + \tfrac{1}{2}n^{-1}\text{tr}[\Omega_1^{-1}d\Omega_1\Omega_1^{-1}(Y - X\Pi_1)'(Y - X\Pi_1)] \\ &\quad + n^{-1}\text{tr}[\Omega_1^{-1}(d\Pi_1)'X'(Y - X\Pi_1)] \\ &= -\tfrac{1}{2}\text{tr}[d\Omega_1(\Omega_1^{-1} - \Omega_1^{-1}n^{-1}(Y - X\Pi_1)'(Y - X\Pi_1)\Omega_1^{-1})] \\ &\quad + n^{-1}\text{tr}[(d\Pi_1)'X'(Y - X\Pi_1)\Omega_1^{-1}] \\ &= -\tfrac{1}{2}(dv(\Omega_1))'\mathbf{D}_m(\Omega_1^{-1} \otimes \Omega_1^{-1})\mathbf{D}_m'v[\Omega_1 - n^{-1}(Y-X\Pi_1)'(Y-X\Pi_1)] \\ &\quad + n^{-1}(d\pi_1)'(\Omega_1^{-1} \otimes \mathbf{I}_{k_1})\text{vec}[X'(Y - X\Pi_1)], \end{aligned}$$

<3.3.3.1>

where

π_1 = vec Π_1

and

$\mathbf{D}_m'v(\Omega_1)$ = vec Ω_1.

To find

$d\phi_2(\delta)$,

an argument based on the chain rule is required, for

$$\Pi_1 = -B_1A_1^{-1},$$

a function of

$$g_1 = \mathrm{vec}\begin{bmatrix} A_1 \\ B_1 \end{bmatrix}$$

and in turn,

$$g_1 = K\delta + k.$$

Thus,

$$\begin{aligned} d\Pi_1 &= -d(B_1A_1^{-1}) \\ &= -dB_1A_1^{-1} - B_1dA_1^{-1} \\ &= -dB_1A_1^{-1} + B_1A_1^{-1}dA_1A_1^{-1} \\ &= -[-B_1A_1^{-1} : I_{k_1}]\begin{bmatrix} dA_1 \\ dB_1 \end{bmatrix}A_1^{-1} \\ &= -(\Pi_1 : I_{k_1})dC_1A_1^{-1} \\ &= -Q_1dC_1A_1^{-1}, \end{aligned} \qquad \langle 3.3.3.2\rangle$$

so that

$$\begin{aligned} d\mathrm{vec}\,\Pi_1 &= -(A_1^{-1\prime} \otimes Q_1)\mathrm{vec}dC_1 \\ &= -(A_1^{-1\prime} \otimes Q_1)dg_1 \\ &= -(A_1^{-1\prime} \otimes Q_1)Kd\delta. \end{aligned}$$

That is,

$$d\phi_2(\delta) = -(A_1^{-1\prime} \otimes Q_1)Kd\delta. \qquad \langle 3.3.3.3\rangle$$

<u>3.3.4.</u> Thus, using the definition of $\mathcal{L}_1$, equation <3.3.2.2>, one can establish that

$$\begin{aligned} d\mathcal{L}_1 = &-\tfrac{1}{2}(dv(\Omega_1))'\mathbf{D}_m(\Omega_1^{-1} \otimes \Omega_1^{-1})\mathbf{D}_m'\{v(\Omega_1) - n^{-1}[(Y-X\Pi_1)'(Y-X\Pi_1)]\} \\ &+ n^{-1}(d\pi_1)'(\Omega_1^{-1} \otimes X')\mathrm{vec}(Y - X\Pi_1) + d\mu_2'(\pi_1 - \phi_2(\delta)) \\ &+ \mu_2'd\pi_1 + \mu_2'(A_1^{-1\prime} \otimes Q_1)Kd\delta, \end{aligned}$$

and this can be arranged in matrix form as

$$d\mathcal{L}_1 = (dv'(\Omega_1) : d\pi_1' : d\delta' : d\mu_2') \times \begin{bmatrix} -\frac{1}{2}\mathbf{D}_m(\Omega_1^{-1}\otimes\Omega_1^{-1})\mathbf{D}_m' v[\Omega_1 - n^{-1}(Y-X\Pi_1)'(Y-X\Pi_1)] \\ n^{-1}(\Omega_1^{-1} \otimes X')\mathrm{vec}(Y - X\Pi_1) + \mu_2 \\ K'(A_1^{-1} \otimes \Box_1')\mu_2 \\ \pi_1 - \phi_2(\delta) \end{bmatrix}.$$

<3.3.4.1>

The maximum likelihood estimators, $\tilde{\Pi}_1$, $\tilde{\Omega}_1$, $\tilde{\delta}$ and $\tilde{\mu}_2$ must therefore satisfy the equations

$$0 = -\tfrac{1}{2}\mathbf{D}_m(\tilde{\Omega}_1^{-1} \otimes \tilde{\Omega}_1^{-1})\mathbf{D}_m' v[\tilde{\Omega}_1 - n^{-1}(Y-X\tilde{\Pi}_1)'(Y-X\tilde{\Pi}_1)] \qquad \text{<3.3.4.2>}$$

$$0 = n^{-1}(\tilde{\Omega}_1^{-1} \otimes X')\mathrm{vec}(Y - X\tilde{\Pi}_1) + \tilde{\mu}_2 \qquad \text{<3.3.4.3>}$$

$$0 = K'(\tilde{A}_1^{-1} \otimes \tilde{\Box}_1')\tilde{\mu}_2 \qquad \text{<3.3.4.4>}$$

$$0 = \tilde{\pi}_1 - \phi_2(\tilde{\delta}). \qquad \text{<3.3.4.5>}$$

An expression for $\tilde{\delta}$ can be obtained as follows: substituting for $\tilde{\mu}_2$ from equation <3.3.4.3> into equation <3.3.4.4>, one obtains

$$\begin{aligned} 0 &= -n^{-1}K'(\tilde{A}_1^{-1} \otimes \tilde{\Box}_1')(\tilde{\Omega}_1^{-1} \otimes X')\mathrm{vec}(Y - X\tilde{\Pi}_1) \\ &= -n^{-1}K'(\tilde{A}_1^{-1}\tilde{\Omega}_1^{-1} \otimes \tilde{\Box}_1'X')(\tilde{A}_1^{-1}\tilde{A}_1' \otimes I_n)\mathrm{vec}(Y - X\tilde{\Pi}_1) \\ &= -n^{-1}K'(\tilde{A}_1^{-1}\tilde{\Omega}_1^{-1}\tilde{A}_1^{-1\prime} \otimes \tilde{\Box}_1'X')\mathrm{vec}((Y - X\tilde{\Pi}_1)\tilde{A}_1) \\ &= -n^{-1}K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Box}_1'X')\mathrm{vec}\, Z_1\tilde{C}_1 \\ &= -n^{-1}K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Box}_1'X'Z_1)\tilde{g}_1 \\ &= -n^{-1}K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Box}_1'X'Z_1)(K\tilde{\delta} + k), \end{aligned} \qquad \text{<3.3.4.6>}$$

since

$$(Y - X\tilde{\Pi}_1)\tilde{A}_1 = Y\tilde{A}_1 + X\tilde{B}_1 = Z_1\tilde{C}_1$$

by equations <3.1.2.3> and <3.1.3.3>, and

$$\Sigma_1 = A_1'\Omega_1 A_1$$

by equation <3.1.2.1>.

Thus, $\tilde{\delta}$ satisfies

$$K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1' X' Z_1) K \tilde{\delta} = - K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1' X' Z_1) k; \qquad \text{<3.3.4.7>}$$

these equations will be called the full information maximum likelihood (FIML) normal equations for δ, and can be regarded as a version of the equations given by Hendry [1976], but where the structural parameter restrictions have been made explicit. The other estimators satisfy

$$\tilde{\Omega}_1 = n^{-1}(Y - X\tilde{\Pi}_1)'(Y - X\tilde{\Pi}_1), \qquad \text{<3.3.4.8>}$$

$$\tilde{\mu}_2 = - n^{-1}(\tilde{\Omega}_1^{-1} \otimes X')\mathrm{vec}(Y - X\tilde{\Pi}_1)$$

$$\tilde{\pi}_1 = \phi_2(\tilde{\delta}).$$

Whilst these four expressions cannot be described as "solutions", since they are interdependent, none the less, they are useful for a number of purposes, as will be seen.

3.3.5. The parameters $v(\Omega_0)$, Π_0 and γ of the null hypothesis model, summarily described by equations <3.1.2.2>, <3.1.2.7>, and <3.1.3.1>,

$$YA_1 + XB_1 = Z_1C_1 = U_1,$$

$$g_1 = K\delta + k,$$

$$\delta = L\gamma + r,$$

rather than the "full model" described in subsection 3.1.3., can be estimated by maximising

$$n^{-1} l_n(y;\phi)$$

subject now to two sets of restrictions,

$$\pi_1 = \phi_2(\delta) \text{ and } \delta = L\gamma + r.$$

Let the corresponding Lagrange multipliers be ξ_2 and ζ_2: the partition of the Lagrange multipliers

$$\xi' = (\xi_1' : \xi_2'), \quad \zeta' = (\zeta_1' : \zeta_2')$$

corresponds to the partition of μ into μ_1 ($\equiv 0$) and μ_2, with ξ_1 and ζ_1 having the same dimension as μ_1, ξ_2 the same dimension (mk_1) as μ_2, and ζ_2 being $q_1 \times 1$.

The Lagrangean is

$$\mathcal{L}_0 = n^{-1}l_n(y;\phi) + \xi_2'(\pi_1 - \phi_2(\delta)) + \zeta_1'(\delta - L\gamma - r),$$

and the differential is easily obtained, using equation <3.3.4.1>, as

$$d\mathcal{L}_0 = (dv'(\Omega_1) : d\pi_1' : d\delta' : d\xi_2' : d\gamma' : d\zeta_2') \times \begin{bmatrix} -\tfrac{1}{2}\mathbf{D}_m(\Omega_1^{-1} \otimes \Omega_1^{-1})\mathbf{D}_m' v\{\Omega_1 - n^{-1}(Y-X\Pi_1)'(Y-X\Pi_1)\} \\ n^{-1}(\Omega_1^{-1} \otimes X')\mathrm{vec}(Y - X\Pi_1) + \xi_2 \\ K'(A_1^{-1} \otimes \mathbf{B}_1')\xi_2 + \zeta_2 \\ \pi_1 - \phi_2(\delta) \\ -L'\zeta_2 \\ \delta - L\gamma - r \end{bmatrix}.$$

Denote the maximum likelihood estimators of the parameters of the null hypothesis model by

$$\tilde{\Omega}_0,\ \tilde{\Pi}_0,\ \tilde{\gamma}$$

and

$$\tilde{\delta}_0 = L\tilde{\gamma} + r,$$

to distinguish it from the estimator $\tilde{\delta}$ from the alternative hypothesis model. The estimators of the quantities A_0, C_0, Σ_0 and $\mathbf{B}_0$ are denoted

$$\tilde{A}_0,\ \tilde{C}_0,\ \tilde{\Sigma}_0,\ \tilde{\mathbf{B}}_0,$$

respectively, with estimated Lagrange multipliers $\tilde{\xi}_2$, $\tilde{\zeta}_2$. These quantities will satisfy the equations

$$0 = -\tfrac{1}{2}\mathbf{D}_m(\tilde{\Omega}_0^{-1} \otimes \tilde{\Omega}_0^{-1})\mathbf{D}_m' v\{\tilde{\Omega}_0 - n^{-1}(Y-X\tilde{\Pi}_0)'(Y-X\tilde{\Pi}_0)\} \qquad \text{<3.3.5.1>}$$

$$0 = n^{-1}(\tilde{\Omega}_0^{-1} \otimes X')\mathrm{vec}(Y - X\tilde{\Pi}_0) + \tilde{\xi}_2 \qquad \text{<3.3.5.2>}$$

$$0 = K'(\tilde{A}_0^{-1} \otimes \tilde{B}_0')\tilde{\xi}_2 + \tilde{\zeta}_2 \qquad \text{<3.3.5.3>}$$

$$0 = \tilde{\pi}_0 - \phi_2(\tilde{\delta}_0) \qquad \text{<3.3.5.4>}$$

$$0 = - L'\tilde{\zeta}_2 \qquad \text{<3.3.5.5>}$$

$$0 = \tilde{\delta}_0 - L\tilde{\gamma} - r\,. \qquad \text{<3.3.5.6>}$$

It will then follow that

$$- L'K'(\tilde{A}_0^{-1} \otimes \tilde{B}_0')\tilde{\xi}_2 = 0$$

or, following the argument leading up to equation <3.3.4.7>,

$$\begin{aligned} 0 &= - L'K'(\tilde{A}_0^{-1}\tilde{\Omega}_0^{-1}\tilde{A}_0^{-1\prime} \otimes \tilde{B}_0'X')\mathrm{vec}\ Z_1\tilde{C}_0 \\ &= - L'K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)\tilde{g}_0 \\ &= - L'K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)(K\tilde{\delta}_0 + k) \\ &= - L'K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)(K(L\tilde{\gamma} + r) + k). \end{aligned}$$

That is,

$$L'K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)KL\tilde{\gamma} = - L'K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)(Kr + k),$$

or using equations <3.1.4.1> and <3.1.4.2>,

$$H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)H\tilde{\gamma} = - H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)h. \qquad \text{<3.3.5.7>}$$

The corresponding estimators of δ, Π_0, Ω_0 and the Lagrange multipliers are

$$\tilde{\Omega}_0 = n^{-1}(Y - X\tilde{\Pi}_0)'(Y - X\tilde{\Pi}_0) \qquad \text{<3.3.5.8>}$$

$$\tilde{\delta}_0 = L\tilde{\gamma} + r,$$

$$\tilde{\pi}_0 = \phi_2(\tilde{\delta}_0) \qquad \text{<3.3.5.9>}$$

$$\tilde{\xi}_2 = - n^{-1}(\tilde{\Omega}_0^{-1} \otimes X')\mathrm{vec}(Y - X\tilde{\Pi}_0),$$

$$\tilde{\zeta}_2 = - K'(\tilde{A}_0^{-1} \otimes \tilde{B}_0')\tilde{\xi}_2.$$

<u>3.3.6.</u> It is useful at this stage to recall the parametric structure into which these "null hypothesis" and "alternative hypothesis" simultaneous equations models have been embedded: the log likelihood function depends on the parameter vector

θ, described as ϕ in the alternative hypothesis model, and which contains the reduced form parameters

$$\theta = \begin{bmatrix} v(\Omega_0) \\ \pi_0 \end{bmatrix}, \quad \phi = \begin{bmatrix} v(\Omega_1) \\ \pi_1 \end{bmatrix};$$

θ is a function of a free parameter vector α,

$$\alpha = \begin{bmatrix} v(\Omega_0) \\ \gamma \end{bmatrix},$$

whilst ϕ depends on the vector β:

$$\beta = \begin{bmatrix} v(\Omega_1) \\ \delta \end{bmatrix}.$$

The alternative hypothesis is defined by equation <2.1.1.1>,

$$H_1: \quad \theta = \phi(\beta),$$

whilst the null hypothesis may be described either by equation <2.1.1.3>,

$$H_0: \quad \theta = \theta(\alpha)$$

or by equation <2.1.1.2>,

$$H_0: \quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha),$$

where the function $\lambda(\alpha)$ is here defined as

$$\lambda(\alpha) = \begin{bmatrix} v(\Omega_0) \\ L\gamma + r \end{bmatrix}.$$

Partitioning the functions $\theta(.)$ and $\lambda(.)$,

$$\theta(.) = \begin{bmatrix} \theta_1(v(\Omega_0)) \\ \theta_2(\gamma) \end{bmatrix}, \; \lambda(.) = \begin{bmatrix} \lambda_1(v(\Omega_0)) \\ \lambda_2(\gamma) \end{bmatrix},$$

one can see that the representation

$$\theta(\alpha) = \phi[\lambda(\alpha)]$$

is equivalent to

$\phi_1[\lambda_1(v(\Omega_0))] = v(\Omega_0)$

and

$\phi_2[\lambda_2(\gamma)] = \theta_2(\gamma)$

so that $\lambda_1(.)$ is the identity function, since $\phi_1(.)$ is the identity function - see subsection 3.3.2. .

Bearing this in mind, it should be clear that estimating the parameters of the null hypothesis model

$(I_m \otimes Z_1)g_0 = u_0$

$g_0 = H\gamma + h$

by maximising the log-likelihood function having the same form as equation <3.3.2.1>,

$$n^{-1}l_n(y;\theta) = ms - \tfrac{1}{2}\log\det\Omega_0 - \tfrac{1}{2}n^{-1}\mathrm{tr}[\Omega_0^{-1}(Y - X\Pi_0)'(Y - X\Pi_0)] \qquad \langle 3.3.6.1\rangle$$

subject to

$\pi_0 = \theta_2(\gamma)$,

with associated Lagrange multiplier τ_2, will produce the same estimators of γ, π_0 and Ω_0 as those given by equations <3.3.5.7>-<3.3.5.9>, but where

$\tilde{\pi}_0 = \phi_2(\tilde{\delta}_0) = \phi_2(L\tilde{\gamma} + r) \equiv \theta_2(\tilde{\gamma})$.

Similarly, the Lagrange multiplier $\tilde{\tau}_2$ will be

$$\tilde{\tau}_2 = -\,n^{-1}(\tilde{\Omega}_0^{-1} \otimes X')\mathrm{vec}(Y - X\tilde{\Pi}_0)$$
$$\equiv \tilde{\xi}_2.$$

Which of these "long" or "short" approaches to the estimation of the null hypothesis model is used at a particular point in this work will depend simply on which is the most convenient for the purpose at hand.

3.3.7. It will be convenient for later sections of this Chapter to establish the nature of the sampling error of the maximum likelihood estimator $\tilde{\gamma}$ about the true value under the null hypothesis, γ^0.

The true value of g_0 under the null hypothesis is

$$g^0 = H\gamma^0 + h,$$

so that

$$(I_m \otimes Z_1)g^0 = u_0 = (I_m \otimes Z_1)(H\gamma^0 + h)$$

and

$$(I_m \otimes Z_1)h = u_0 - (I_m \otimes Z_1)H\gamma^0.$$

Substituting for $(I_m \otimes Z_1)h$ in the expression for $\tilde{\gamma}$ derived from equation <3.3.5.7>, one obtains

$$\begin{aligned}
\tilde{\gamma} &= - [H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'Z_1)H]^{-1}H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'Z_1)h \\
&= - [H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'Z_1)H]^{-1}H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X')(u_0 - (I_m \otimes Z_1)H\gamma^0) \\
&= \gamma^0 - [H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'Z_1)H]^{-1}H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X')u_0 \\
&= \gamma^0 - [H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'Z_1)H]^{-1}H'(\tilde{A}_0^{-1}\tilde{\Omega}_0^{-1} \otimes \tilde{\Pi}_0'X')v_0, \qquad \text{<3.3.7.1>}
\end{aligned}$$

since

$$V_0 = U_0A_0^{-1}.$$

One could in fact use the strong consistency of $\tilde{\Sigma}_0$ and $\tilde{\Pi}_0$ to obtain the limiting distribution of $\tilde{\gamma}$ directly from these latter two expressions, rather than as a deduction from the general maximum likelihood results, which is the approach taken in the next section.

3.4 Asymptotic Properties of the Maximum Likelihood Estimators.

3.4.1. For the general arguments of Chapter 2, a number of critical assumptions are required to establish the strong consistency of the maximum likelihood estimators of θ, α and ϕ, β under the null hypothesis: that the scaled log-likelihood function

$n^{-1}l_n(y;\theta)$

converges almost surely and uniformly in θ to a function

$l(\theta^0;\theta)$

having a unique maximum at θ^0, that the parameter spaces $\mathcal{A}$ and $\mathcal{B}$ of α and β respectively are compact, and that α and β are uniquely identified at the true value θ^0.

In the case of the simultaneous equations model, one can find the limit function $l(\theta^0;\theta)$ explicitly: the log-likelihood function is given by equation <3.3.6.1> as

$$n^{-1}l_n(y;\theta) = ms - \tfrac{1}{2}\log\det\Omega_0 - \tfrac{1}{2}n^{-1}\mathrm{tr}[\Omega_0^{-1}(Y - X\Pi_0)'(Y - X\Pi_0)]. \quad <3.4.1.1>$$

The observed values $y_1,\ldots,y_n$ are generated by the true parameter values Π^0, Ω^0 according to

$y_t = \Pi^{0\prime}x_t + v_{0t}, \quad t = 1,\ldots,n,$

with the independent and identically distributed random vectors v_{0t}, $t = 1,\ldots,n$ having mean 0 and covariance matrix Ω^0; the observation matrix form is

$Y = X\Pi^0 + V_0.$

Thus,

$$n^{-1}l_n(y;\theta) = ms - \tfrac{1}{2}\log\det\Omega_0 - \tfrac{1}{2}n^{-1}\mathrm{tr}\{\Omega_0^{-1}[V_0 - X(\Pi_0 - \Pi^0)]' \times [V_0 - X(\Pi_0 - \Pi^0)]\}$$

$$= ms - \tfrac{1}{2}\log\det\Omega_0 - \tfrac{1}{2}n^{-1}\mathrm{tr}\{\Omega_0^{-1}V_0'V_0 - 2V_0'X(\Pi_0 - \Pi^0) + (\Pi_0 - \Pi^0)'X'X(\Pi_0 - \Pi^0)\}.$$

Application of Khintchine's Strong Law of Large Numbers (see for example, Rao [1973, p215], who refers to this as Kolmogorov's "Theorem 2") then shows that

$$n^{-1}\sum_{t=1}^{n} v_{0t}v_{0t}' = n^{-1}V_0'V_0 \xrightarrow{a.s.} \Omega^0.$$

Next, writing

$$V_0 = (v_{0.1},\ldots,v_{0.m}),$$

the columns of the matrix

$$n^{-1}X'V_0$$

may be denoted

$$n^{-1}X'v_{0.i}, \qquad i = 1,\ldots,m,$$

and equal

$$n^{-1}\sum_{t=1}^{n} x_t v_{0.it}, \qquad i = 1,\ldots,m,$$

where the elements of $v_{0.i}$ are $v_{0.it}$, $t = 1,\ldots,n$. The boundedness assumption on the elements of the vectors x_t (see subsection 3.1.2.) then ensures that each one of the columns of $n^{-1}X'V_0$ satisfies the conditions of Kolmogorov's Strong Law of Large Numbers (referred to by Rao [1973,p114] as Kolmogorov's Theorem 1), so that

$$n^{-1}X'V_0 \xrightarrow{a.s.} 0.$$

Finally, it will be assumed, throughout this thesis, that

$$n^{-1}X'X \rightarrow M_X,$$

a positive definite matrix.

Assembling all these results, it follows that

$$n^{-1}l_n(y;\theta) \xrightarrow{a.s.} ms - \tfrac{1}{2}\log\det\Omega_0 - \tfrac{1}{2}tr[\Omega_0^{-1}\Omega^0 + \Omega_0^{-1}(\Pi_0-\Pi^0)'M_x(\Pi_0-\Pi^0)]$$

$$= l(\theta^0;\theta).$$

One can show, using the argument given by Rao [1973, p531] that

$$l(\theta^0;\theta^0) - l(\theta^0;\theta) \geq 0$$

if and only if $\Pi_0 = \Pi^0$, $\Omega_0 = \Omega^0$, which establishes the correctness of the unique maximum assumption for the simultaneous equations case.

Of the two remaining assumptions, the unique identification assumption has already been discussed in section 3.2., and the compactness of the parameter spaces of α and β is in general non-verifiable.

Thus, it is possible to claim that the maximum likelihood estimators of the alternative and null hypothesis simultaneous equations models,

$\tilde{\Pi}_1$, $\tilde{\Omega}_1$, $\tilde{\delta}$

and

$\tilde{\Pi}_0$, $\tilde{\Omega}_0$, $\tilde{\gamma}$, $\tilde{\delta}_0$

converge almost surely to the true values

Π^0, Ω^0, γ^0 and $\delta_0^0 = L\gamma^0 + r$,

given the truth of the null hypothesis. Almost sure convergence to zero for the corresponding Lagrange multipliers will follow using exactly the same Strong Laws of Large Numbers.

3.4.2. The basis of the limit normal distribution obtained for the maximum likelihood estimators in section 2.4. is the assumption that the "score" vector $D_\theta l_n(y;\theta^0)$ satisfies a Central Limit Theorem, so that

$$n^{-\frac{1}{2}}D_\theta l_n(y;\theta^0) \xrightarrow{d} w \sim N(0, I(\theta^0)).$$

In the simultaneous equations model, the score vector is, from the differential $n^{-1}dl_n$ of equation <3.3.3.1> (and adjusting the notation to suit the null hypothesis model),

$$n^{-1}D_\theta l_n(y;\theta^0) = \begin{bmatrix} -\frac{1}{2}\mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m' v[\Omega^0 - n^{-1}(Y-X\Pi^0)'(Y-X\Pi^0)] \\ n^{-1}(\Omega^{0-1} \otimes X')\mathrm{vec}(Y - X\Pi^0) \end{bmatrix} \qquad \text{<3.4.2.1>}$$

$$= \begin{bmatrix} -\frac{1}{2}\mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m' v[(\Omega^0 - n^{-1}V_0'V_0] \\ n^{-1}(\Omega^{0-1} \otimes X')\mathrm{vec}\ V_0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2}\mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m' n^{-1} v[\sum_{t=1}^{n}(v_{0t}v_{0t}' - \Omega^0)] \\ (\Omega^{0-1} \otimes I_{k_1})n^{-1}\sum_{t=1}^{n}\mathrm{vec}\ (x_t v_{0t}') \end{bmatrix}.$$

A Central Limit Theorem for independently and identically distributed random vectors can be applied to the first subvector, and one for independent but not identically distributed random vectors to the second; jointly, then, it will be possible to apply a Lindeberg-Feller or a Liapunov Central Limit Theorem (with their appropriate assumptions) to an arbitrary linear combination of the elements of $n^{-\frac{1}{2}}D_\theta l_n(y;\theta^0)$. It seems inappropriate to labour through the tedious details of verification of the conditions required by these theorems.

The covariance matrix (under the normality assumption) of this random vector can be established from the appropriate version of equation <1.6.7.2.> as

$$\mathrm{var}_{\theta^0} \sum_{t=1}^{n} \begin{bmatrix} v(v_{ot}v_{ot}') \\ \mathrm{vec}(x_t v_{ot}') \end{bmatrix}$$

$$= \sum_{t=1}^{n} \begin{bmatrix} 2\mathbf{L}_m\mathbf{S}_m(\Omega^0 \otimes \Omega^0)\mathbf{S}_m'\mathbf{L}_m' & \vdots & 0 \\ 0 & \vdots & (\Omega^0 \otimes x_t x_t') \end{bmatrix}.$$

To obtain the desired covariance matrix, note that

$$\sum_{t=1}^{n} x_t x_t' = X'X,$$

and

$$\mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m'\mathbf{L}_m\mathbf{S}_m(\Omega^0 \otimes \Omega^0)\mathbf{S}_m'\mathbf{L}_m'\mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m'$$

$$= \mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{S}_m(\Omega^0 \otimes \Omega^0)\mathbf{S}_m'(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m'$$

$$= \mathbf{D}_m(\Omega^{0-1}\Omega^0\Omega^{0-1} \otimes \Omega^{0-1}\Omega^0\Omega^{0-1})\mathbf{D}_m'$$

$$= \mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m',$$

using the results <1.6.3.6>, <1.6.3.12>, <1.6.3.13>.

Then, using the definition in equation <2.4.1.3>,

$$\mathbf{I}_n(\theta^0) = n^{-1}[\![E_\theta D_\theta l_n(y;\theta^0) D_\theta l_n'(y;\theta^0)]\!]_{\theta^0}$$

$$= n^{-1}\begin{bmatrix} \tfrac{1}{2}n\mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m' & \vdots & 0 \\ 0 & \vdots & (\Omega^{0-1} \otimes X'X) \end{bmatrix}, \qquad \langle 3.4.2.2\rangle$$

and it is clear, given the assumption that

$$n^{-1}X'X \rightarrow M_x,$$

that

$$\mathbf{I}_n(\theta) = \begin{bmatrix} \tfrac{1}{2}\mathbf{D}_m(\Omega_0^{-1} \otimes \Omega_0^{-1})\mathbf{D}_m' & \vdots & 0 \\ 0 & \vdots & (\Omega_0^{-1} \otimes n^{-1}X'X) \end{bmatrix} \qquad \langle 3.4.2.3\rangle$$

converges uniformly in θ to the matrix

$$\mathbf{I}(\theta) = \begin{bmatrix} \tfrac{1}{2}\mathbf{D}_m(\Omega_0^{-1} \otimes \Omega_0^{-1})\mathbf{D}_m' & \vdots & 0 \\ 0 & \vdots & (\Omega_0^{-1} \otimes M_x) \end{bmatrix} \qquad \langle 3.4.2.4\rangle$$

with inverse matrix (using equation <1.6.3.15>)

$$I^{-1}(\theta) = \begin{bmatrix} 2L_mS_m(\Omega_0 \otimes \Omega_0)S_m'L_m' : & 0 \\ 0 & :(\Omega_0 \otimes M_X^{-1}) \end{bmatrix}.$$

Overall, then, under the assumptions made for the simultaneous equations model,

$n^{-1/2}D_\theta l_n(y;\theta^0) \xrightarrow{d} w \sim N(0, I(\theta^0))$,

with $I(\theta^0)$ defined by equation <3.4.2.4> evaluated at θ^0.

<u>3.4.3.</u> One can now carry over the limit normal distribution results of section 2.4. to the simultaneous equations model: it is perhaps appropriate to recall the results of that section. The statistical model for the observed random vectors $y_1,\dots,y_n$ is subsumed within the log-likelihood function

$n^{-1}l_n(y;\theta)$;

under the alternative hypothesis,

H_1: $\theta = \phi(\beta)$,

whilst under the null hypothesis, β is further restricted, so that

H_0: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$,

or

H_0: $\theta = \phi[\lambda(\alpha)] = \theta(\alpha)$.

The maximum likelihood estimators of θ and β from the alternative hypothesis model, denoted $\tilde{\phi}$, $\tilde{\beta}$, with the corresponding Lagrange multiplier $\tilde{\mu}$ (see subsection 2.4.1.) are arranged in the vector

$\tilde{\psi}_1' = (\tilde{\phi}', \tilde{\beta}', \tilde{\mu}')$: <3.4.3.1>

if the vector of true values under the null hypothesis is

$$\psi_1^{0\prime} = (\theta^{0\prime},\ \beta^{0\prime},\ 0), \qquad \text{<3.4.3.2>}$$

then

$$n^{1/2}(\tilde{\psi}_1 - \psi_1^0) \overset{a}{\approx} N(0,\ \Psi(\tilde{\psi}_1;\psi_1^0)),$$

where, from equation <2.4.1.12>,

$$\Psi(\tilde{\psi}_1;\psi_1^0) = \begin{bmatrix} \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi' & : & \Phi(\Phi'\mathbf{I}\Phi)^{-1} & : & 0 \\ (\Phi'\mathbf{I}\Phi)^{-1}\Phi' & : & (\Phi'\mathbf{I}\Phi)^{-1} & : & 0 \\ 0 & : & 0 & : & \mathbf{IP}_\Phi \end{bmatrix} \qquad \text{<3.4.3.3>}$$

In this expression,

$\Phi = \Phi(\beta^0) = D_\beta\phi(\beta^0)$, $\mathbf{I} = \mathbf{I}(\theta^0)$,

and $\mathbf{P}_\Phi$ is defined by equation <2.4.1.9>:

$$\begin{aligned} \mathbf{P}_\Phi &= \mathbf{I}_{s_0} - \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi'\mathbf{I} \\ &= \mathbf{I}_{s_0} - \Phi(\beta^0)(\Phi'(\beta^0)\mathbf{I}(\theta^0)\Phi(\beta^0))^{-1}\Phi'(\beta^0)\mathbf{I}(\theta^0) \\ &= \mathbf{P}_\Phi(\beta^0). \end{aligned}$$

For the null hypothesis model, the maximum likelihood estimators are denoted $\tilde{\theta}$, $\tilde{\beta}_0$, $\tilde{\alpha}$ with corresponding Lagrange multipliers $\tilde{\xi}$ and $\tilde{\zeta}$ (see subsection 2.4.2.) arranged in the vector

$$\tilde{\psi}_0' = (\tilde{\theta}',\ \tilde{\beta}_0',\ \tilde{\xi}',\ \tilde{\alpha}',\ \tilde{\zeta}')$$

with corresponding "true value" vector

$$\psi_0^{0\prime} = (\theta^{0\prime},\ \beta^{0\prime},\ 0,\ \alpha^{0\prime},\ 0), \qquad \text{<3.4.3.4>}$$

$$n^{1/2}(\tilde{\psi}_0 - \psi_0^0) \overset{a}{\approx} N(0,\ \Psi(\tilde{\psi}_0;\psi_0^0));$$

to express the matrix $\Psi(\tilde{\psi}_0;\psi_0^0)$ succinctly, let $\mathbf{A}$ be the matrix given in equation <2.4.2.2>,

$$A = \begin{bmatrix} B(B'IB)^{-1}B' \\ \Lambda(B'IB)^{-1}B' \\ -P' \\ (B'IB)^{-1}B' \\ -\Phi'P' \end{bmatrix}, \qquad \langle 3.4.3.5\rangle$$

where

$B = B(\alpha^0) = D_\alpha\theta(\alpha^0)$,

$\Lambda = \Lambda(\alpha^0) = D_\alpha\lambda(\alpha^0)$,

$B(\alpha^0) = \Phi(\beta^0)\Lambda(\alpha^0)$

and P is defined by equation <2.4.2.3>,

$$\begin{aligned} P &= I_{s_0} - B(B'IB)^{-1}B'I \\ &= I_{s_0} - B(\alpha^0)(B'(\alpha^0)I(\theta^0)B(\alpha^0))^{-1}B'(\alpha^0)I(\theta^0) \\ &= P(\alpha^0). \end{aligned}$$

Then,

$$\Psi(\tilde{\psi}_0;\psi_0^0) = AI(\theta^0)A'. \qquad \langle 3.4.3.6\rangle$$

For the simultaneous equations model, the vector functions $\phi = \phi(\beta)$, $\beta = \lambda(\alpha)$, $\theta = \theta(\alpha)$ are

$$\begin{bmatrix} v(\Omega_1) \\ \pi_1 \end{bmatrix} = \begin{bmatrix} v(\Omega^0) \\ \phi_2(\delta) \end{bmatrix},$$

$$\begin{bmatrix} v(\Omega_1) \\ \phi_2(\delta) \end{bmatrix} = \begin{bmatrix} v(\Omega_0) \\ L\gamma + r \end{bmatrix},$$

$$\begin{bmatrix} v(\Omega_0) \\ \pi_0 \end{bmatrix} = \begin{bmatrix} v(\Omega_0) \\ \theta_2(\gamma) \end{bmatrix},$$

respectively; the derivative matrices are

$$\Phi = \begin{bmatrix} I_{\frac{1}{2}m(m+1)} & : & 0 \\ 0 & : & -(A_1^{-1\prime} \otimes \Omega_1)K \end{bmatrix} \qquad \langle 3.4.3.7\rangle$$

(using <3.3.3.3>);

$$\Lambda = \begin{bmatrix} I_{\frac{1}{2}m(m+1)} & : & 0 \\ 0 & : & L \end{bmatrix},$$

$$\Theta = \begin{bmatrix} I_{\frac{1}{2}m(m+1)} & : & 0 \\ 0 & : & -(A_0^{-1\prime} \otimes \Omega_0)H \end{bmatrix}. \qquad \langle 3.4.3.8\rangle$$

With this information, and the limiting information matrix of equation <3.4.2.4>, the quantities in $\Psi(\tilde{\psi}_1;\psi_1^0)$ and $\Psi(\tilde{\psi}_0;\psi_0^0)$ can be constructed. Only the diagonal submatrices of these latter two matrices which are employed later in the thesis will be given explicitly.

The limiting covariance matrix of $\tilde{\delta}$ from the alternative hypothesis model is given by the second diagonal submatrix of

$$(\Phi' I \Phi)^{-1} = \begin{bmatrix} \tfrac{1}{2}\mathbf{D}_m(\Omega^{0-1} \otimes \Omega^{0-1})\mathbf{D}_m' : & 0 \\ 0 & : K'(A^{0-1}\Omega^{0-1}A^{0-1\prime} \otimes \Omega^{0\prime}M_X\Omega^0)K \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 2\mathbf{L}_m\mathbf{S}_m(\Omega^0 \otimes \Omega^0)\mathbf{S}_m'\mathbf{L}_m' : & 0 \\ 0 & : [K'(\Sigma^{0-1} \otimes \Omega^{0\prime}M_X\Omega^0)K]^{-1} \end{bmatrix}:$$

<3.4.3.9>

thus,

$$\Psi(\tilde{\delta};\psi_1^0) = [K'(\Sigma^{0-1} \otimes \Omega^{0\prime}M_X\Omega^0)K]^{-1}. \qquad \langle 3.4.3.10\rangle$$

One can similarly obtain from $(\Theta' I \Theta)^{-1}$ the limiting covariance matrix

$$\Psi(\tilde{\gamma};\psi_0^0) = [H'(\Sigma^{0-1} \otimes \Omega^{0\prime}M_X\Omega^0)H]^{-1}. \qquad \langle 3.4.3.11\rangle$$

It is also convenient for the purposes of Chapter 4 to summarise the limit distribution of the maximum likelihood estimator of π_0:

$$n^{\frac{1}{2}}(\tilde{\pi}_0 - \pi^0) \overset{a}{\sim} N(0,\Psi(\tilde{\pi}_0;\theta^0)), \qquad \langle 3.4.3.12\rangle$$

with

$$\Psi(\tilde{\pi}_0;\theta^0) = (A^{0-1\prime} \otimes B^0)H[H'(\Sigma^{0-1} \otimes B^{0\prime}M_xB^0)H]^{-1} \times H'(A^{0-1} \otimes B^{0\prime}). \qquad \text{<3.4.3.13>}$$

The projection matrices $\mathbf{P}$ and $\mathbf{P}_\Phi$ (defined in equations <2.4.2.3> and <2.4.1.9> respectively) can now be constructed:

$$\mathbf{P}_\Phi = \mathbf{I}_{s_0} - \Phi(\Phi'\mathbf{I}\Phi)^{-1}\Phi'\mathbf{I}$$

$$= \begin{bmatrix} 0 & : & 0 \\ 0 & : & P_K \end{bmatrix}$$

where

$$P_K = P_K(\beta^0) = \mathbf{I}_{mk_1} - (A^{0-1\prime} \otimes B^0)K[K'(\Sigma^{0-1} \otimes B^{0\prime}M_xB^0)K]^{-1}K' \times (A^0\Omega^{0-1} \otimes B^{0\prime}M_x),$$

and

$$\mathbf{P} = \mathbf{I}_{s_0} - \Theta(\Theta'\mathbf{I}\Theta)^{-1}\Theta'\mathbf{I}$$

$$= \begin{bmatrix} 0 & : & 0 \\ 0 & : & P \end{bmatrix} \qquad \text{<3.4.3.14>}$$

where

$$P = P(\alpha^0) = \mathbf{I}_{mk_1} - (A^{0-1\prime} \otimes B^0)H[H'(\Sigma^{0-1} \otimes B^{0\prime}M_xB^0)H]^{-1}H' \times (A^{0-1}\Omega^{0-1} \otimes B^{0\prime}M_x). \qquad \text{<3.4.3.15>}$$

The covariance matrices of the limiting normal distributions of the Lagrange multipliers $\tilde{\mu}$, $\tilde{\xi}$ and $\tilde{\zeta}$ then follow from equation <3.4.3.3>,

$$\Psi(\tilde{\mu};\psi_1^0) = \begin{bmatrix} 0 & : & 0 \\ 0 & : & P_K'(\Omega^{0-1} \otimes M_x)P_K \end{bmatrix}$$

so that the non-trivial part is

$$\Psi(\tilde{\mu}_2;\psi_1^0) = P_K'(\Omega^{0-1} \otimes M_x)P_K.$$

Similarly, from equation <3.4.3.6>,

$$\Psi(\tilde{\xi}_2;\psi_0^0) = P'(\Omega^{0-1} \otimes M_x)P,$$

and from equation <3.4.3.5>,

$$\Psi(\tilde{\zeta};\psi_0^0) = \mathbf{\Phi}'\mathbf{P}'\mathbf{I}\mathbf{P}\mathbf{\Phi}$$

$$= \begin{bmatrix} 0 : & 0 \\ 0 : & K'(A^{0-1} \otimes B^{0\prime})P'(\Omega^{0-1} \otimes M_X)P(A^{0-1\prime} \otimes B^0)K \end{bmatrix},$$

the second diagonal block being the relevant one.

3.5. Two Step Estimation.

3.5.1 In this section, the two-step estimation principle discussed in section 2.6. is applied to the simultaneous equations model, partly with a view to investigating the relationship of the maximum likelihood estimators of δ or γ to other estimators that have been proposed, and partly as a basis for the discussion of C-alpha test statistics for tests of over-identifying restrictions in Chapter 5.

Focussing on the null hypothesis model for simplicity, the two-step estimator of α from equation <2.6.4.1> is

$$\hat{\alpha} = \alpha^* + (\mathbf{B}'(\alpha^*)\mathbf{I}_n(\theta^*)\mathbf{B}(\alpha^*))^{-1}\mathbf{B}'(\alpha^*)n^{-1}D_\theta l_n(y;\theta^*),$$

where α^* and

$$\theta^* = \theta(\alpha^*)$$

are appropriate initial estimators. To convert this into an expression for the subvector γ in

$$\alpha = \begin{bmatrix} v(\Omega_0) \\ \gamma \end{bmatrix},$$

note that the structure of the derivative matrix

$$\mathbf{B} = D_\alpha\theta$$

for the simultaneous equations model is given by equation <3.4.3.8>, whilst $\mathbf{I}_n(\theta)$ is given in equation <3.4.2.3> , and the score vector is, by equations <3.3.3.1> or <3.4.2.1>,

$$n^{-1}D_\theta l_n(y;\theta^*) = \begin{bmatrix} -\tfrac{1}{2}\mathbf{D}_m(\Omega_0^{*-1} \otimes \Omega_0^{*-1})\mathbf{D}_m' v[\Omega_0^* - n^{-1}(Y-X\Pi_0^*)'(Y-X\Pi_0^*)] \\ n^{-1}(\Omega_0^{*-1} \otimes X')\mathrm{vec}(Y - X\Pi_0^*) \end{bmatrix}.$$

<3.5.1.1>

It may be assumed that for most initial estimators,

$$\Omega_0^* = n^{-1}(Y - X\Pi_0^*)'(Y - X\Pi_0^*),$$

so that the first term in this score vector is null. The initial estimators γ^*, Π_0^* and Σ_0^* can be used to give an estimator of the matrix $\Psi(\tilde{\gamma};\psi_0^0)$ of equation <3.4.3.11>, denoted $\Psi_n(\tilde{\gamma};\psi_0^*)$ (using the notation of subsection 1.5.4.):

$$\Psi_n(\tilde{\gamma};\psi_0^*) = n[H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H]^{-1}.$$

One can then obtain the two-step estimator of γ as

$$\hat{\gamma} = \gamma^* - [H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H]^{-1}H'(A_0^{*-1}\Omega_0^{*-1} \otimes \Pi_0^{*\prime}X')\mathrm{vec}(Y - X\Pi_0^*) \qquad \text{<3.5.1.2>}$$

$$= \gamma^* - [H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H]^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')\mathrm{vec}[(Y - X\Pi_0^*)A_0^*]$$

$$= \gamma^* - [H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H]^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')\mathrm{vec}\ Z_1C_0^*.$$

The notation A_0^*, C_0^* allows one to observe that

$$\mathrm{vec}\ C_0^* = \mathrm{vec}\begin{bmatrix} A_0^* \\ B_0^* \end{bmatrix} = g_0^* = H\gamma^* + h. \qquad \text{<3.5.1.3>}$$

Since

$$Z_1C_0 = U_0,$$

a residual matrix U_0^* is defined by $Z_1C_0^*$; thus,

$$\mathrm{vec}\ Z_1C_0^* = \mathrm{vec}\ U_0^* = u_0^*, \qquad \text{<3.5.1.4>}$$

and the "update" term in the expression for $\hat{\gamma}$ then depends on these structural form residuals associated with the initial estimator γ^*:

$$\hat{\gamma} = \gamma^* - [H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H]^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')u_0^*. \qquad \text{<3.5.1.5>}$$

The update term can be interpreted as arising from the regression of the residual vector u_0^* on

$$(I_m \otimes X\Pi_0^*)H,$$

in the metric of $\Sigma_0^{*-1} \otimes I_n$; it is also worth noting that this regressor set involves the "fitted values" of the matrix Z_1 based on the implied initial estimator of Π_0, Π_0^*.

There is a corresponding expression for the two-step estimator of δ in the alternative hypothesis model, based on appropriate initial estimators δ^*, Π_1^*, Σ_1^*, namely,

$$\hat{\delta} = \delta^* - [K'(\Sigma_1^{*-1} \otimes \Omega_1^{*\prime}X'X\Omega_1^*)K]^{-1}K'(\Sigma_1^{*-1} \otimes \Omega_1^{*\prime}X')u_1^*,$$

where u_1^* is the structural form residual vector associated with δ^*.

<u>3.5.2.</u> A comparison of the expression <3.5.1.5> above for the two-step estimator $\hat{\gamma}$,

$$\hat{\gamma} = \gamma^* - [H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X'X\Omega_0^*)H]^{-1}H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X')u_0^*$$

with the expression for the FIML estimator $\tilde{\gamma}$ implied by equation <3.3.5.7>,

$$\tilde{\gamma} = -[H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Omega}_0'X'Z_1)H]^{-1}H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Omega}_0'X'Z_1)h,$$

might lead one to ask whether an estimator $\gamma^\dagger$ which is asymptotically equivalent to $\tilde{\gamma}$ could be obtained from the expression

$$\gamma^\dagger = -[H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X'X\Omega_0^*)H]^{-1}H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X'Z_1)h \qquad \text{<3.5.2.1>}$$

using the same initial estimators as the two-step estimator. Such an estimator might be described as a "symmetric" FIML estimator, and could be obtained by a regression of

$$-(I_m \otimes Z_1)h$$

on

$$(I_m \otimes X\Omega_0^*)H$$

with respect to the metric $\Sigma_0^{*-1} \otimes I_n$. This possibility is discussed by Hendry [1976, pp57-58], who notes that if Π_0^* is actually the ordinary least squares (OLS) estimator,

$\hat{\Pi} = (X'X)^{-1}X'Y,$

so that Ω_0^* becomes

$\hat{\Omega} = (X'X)^{-1}X'Z_1,$

$\gamma^\dagger$ and $\tilde{\gamma}$ have the same limiting distribution. If in addition, Σ_0^* is obtained from the two-stage least squares estimator of γ, $\gamma^\dagger$ amounts to the well-known three-stage least squares estimator.

Hendry also observes that

> "the efficiency of $\gamma^\dagger$ does now depend on that of Ω_0^* and will be asymptotically efficient if and only if Ω_0^* is at least as efficient as $\hat{\Omega}$."
> (Hendry [1976,p57])

In fact, this latter statement is only true when Π_0^* and $\hat{\Pi}$ have the same limiting distribution: even when Π_0^* is identical to $\tilde{\Pi}_0$, the FIML estimator associated with $\tilde{\gamma}$, $\gamma^\dagger$ is less efficient asymptotically than $\tilde{\gamma}$.

The proofs of these remarks are a little involved, and will be given in the next subsection. Two "symmetric" FIML estimators having the desired properties will then be discussed in subsection 3.5.4. .

Finally, a small point of notation: the ordinary least squares estimators from the reduced form are not subscripted "0" or "1", simply because they may correspond to either the null or the alternative hypothesis models.

3.5.3. The proof strategy is to obtain directly the formal limit normal distribution of

$n^{1/2}(\gamma^{\dagger} - \gamma^{0})$,

and then make an inspired choice of Π_0^* to establish the existence of a counterexample to Hendry's assertion.

Under the null hypothesis, the true value of γ is γ^0, so that

$$(I_m \otimes Z_1)g^0 = u_0 = (I_m \otimes Z_1)(H\gamma^0 + h)$$

and hence

$$(I_m \otimes Z_1)h = u_0 - (I_m \otimes Z_1)H\gamma^0.$$

Then, from equation <3.5.2.1>,

$$\begin{aligned} H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X'X\Omega_0^*)H\gamma^{\dagger} &= - H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X')(u_0 - (I_m \otimes Z_1)H\gamma^0) \\ &= - H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X')u_0 + H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X'Z_1)H\gamma^0. \end{aligned} \quad \text{<3.5.3.1>}$$

Again, under the null hypothesis, the reduced form is

$$Y = X\Pi^0 + V_0,$$

so that

$$\begin{aligned} Z_1 = (Y : X) &= X(\Pi^0 : I_{k_1}) + (V_0 : O_{nk_1}) \\ &= X\Omega^0 + (V_0 : O_{nk_1}); \end{aligned}$$

using this, the second term in equation <3.5.3.1> can be written as

$$H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X'X\Omega^0)H\gamma^0 + H'(\Sigma_0^{*-1} \otimes \Omega_0^{*\prime}X'(V_0 : O_{nk_1}))H\gamma^0. \quad \text{<3.5.3.2>}$$

In turn, the second term in this expression can be rewritten: the OLS estimator $\hat{\Pi}$ satisfies

$$X'V_0 = X'X(\hat{\Pi} - \Pi^0),$$

under the null hypothesis, so that

$X'(V_0:0_{nk_1}) = X'X(\hat{\Pi} - \Pi^0),$

yielding for the second term in equation <3.5.3.2>,

$H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X(\hat{\Pi} - \Pi^0))H\gamma^0.$

It will be convenient for the argument to follow to define the matrices

$$F_n = n^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H,$$

$$M_n = n^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi^0)H,$$

and the vectors

$$w_{1n} = F_n n^{1/2}\gamma^\dagger - M_n n^{1/2}\gamma^0,$$

$$w_{2n} = F_n n^{1/2}(\gamma^\dagger - \gamma^0),$$

$$w_{3n} = n^{1/2}(F_n - M_n)\gamma^0$$

$$= n^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'Xn^{1/2}(\Pi_0^* - \Pi^0))H\gamma^0.$$

The purpose of the analysis is to find the limit distribution of the scaled sampling error

$$n^{1/2}(\gamma^\dagger - \gamma^0) = F_n^{-1}w_{2n},$$

and it can be seen that w_{2n} satisfies

$$w_{2n} = w_{1n} - w_{3n}.$$

From these definitions, it can be seen that

$$w_{1n} = -\,H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime})n^{-1/2}(I_m \otimes X')u_0$$
$$+\,H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}n^{-1}X'Xn^{1/2}(\hat{\Pi} - \Pi^0))H\gamma^0,$$

so that

$$w_{1n} - w_{3n} = -\,H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime})n^{-1/2}(I_m \otimes X')u_0$$
$$+\,H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}n^{-1}X'X)(I_m \otimes n^{1/2}(\hat{\Pi} - \Pi_0^*))H\gamma^0.$$

Consider the $mk_1 \times 1$ vector

$(I_m \otimes n^{1/2}(\hat{\Pi} - \Pi_0^*))H\gamma^0$:

this can be written as

$$\text{vec}(n^{1/2}(\hat{\mathrm{B}} - \mathrm{B}_0^*)\mathbf{G}) = \text{vec}(n^{1/2}(\hat{\Pi} - \Pi_0^*)\mathbf{G}_1)$$
$$= (\mathbf{G}_1' \otimes I_{k_1})n^{1/2}(\hat{\pi} - \pi_0^*),$$

where $\mathbf{G}$ is a matrix such that

$$\text{vec } \mathbf{G} = H\gamma^0,$$
$$\mathbf{G}' = (\mathbf{G}_1' : \mathbf{G}_2'),$$

with $\mathbf{G}_1$ being $m \times m$, $\mathbf{G}_2$ $k_1 \times m$, and $\hat{\pi} = \text{vec } \hat{\Pi}$.

Thus, one has

$$n^{1/2}(\gamma^\dagger - \gamma^0) = M_n^{-1}[- n^{-1/2}H'(\Sigma_0^{*-1} \otimes \mathrm{B}_0^{*\prime}X')u_0$$
$$+ H'(\Sigma_0^{*-1}\mathbf{G}_1' \otimes \mathrm{B}_0^{*\prime}n^{-1}X'X)\text{vec } n^{1/2}(\hat{\Pi} - \Pi_0^*)].$$
<3.5.3.4>

The first term on the right hand side of this expression has the same limiting distribution as the FIML estimator $\tilde{\gamma}$ under the null hypothesis, whilst the second term will only vanish "in distribution" when

$$n^{1/2}(\tilde{\pi} - \pi_0^*) \xrightarrow{P} 0:$$

that is, when $\tilde{\Pi}$ and Π_0^* have the same limiting distribution and hence the same asymptotic efficiency.

One can make this point rather starkly by choosing Π_0^* to be the FIML estimator $\tilde{\Pi}_0$, which is known (see Dhrymes [1973]) to be more efficient, asymptotically, than the OLS estimator $\hat{\Pi}$ in an overidentified simultaneous equations model. In this case, a limit normal distribution for

$$n^{1/2}\text{vec}(\hat{\Pi} - \tilde{\Pi}_0)$$

can be obtained, since both

$n^{1/2}\text{vec}(\hat{\Pi} - \Pi^0)$ and $n^{1/2}\text{vec}(\tilde{\Pi}_0 - \Pi^0)$

are, asymptotically, linear transformations of

$n^{-1/2}(I_m \otimes X')v_0$.

For the OLS estimator, this follows directly from equation <3.5.3.3>:

$$n^{1/2}\mathrm{vec}(\hat{\Pi} - \Pi^0) = n(I_m \otimes (X'X)^{-1})n^{-1/2}(I_m \otimes X')v_0$$

$$\overset{a}{\approx} (I_m \otimes M_X^{-1})n^{-1/2}(I_m \otimes X')v_0 \qquad \text{<3.5.3.5>}$$

$$\overset{a}{\approx} N(0, \Psi(\hat{\pi};\theta^0)), \qquad \text{<3.5.3.6>}$$

with

$$\Psi(\hat{\pi};\theta^0) = \Omega^0 \otimes M_X^{-1}. \qquad \text{<3.5.3.7>}$$

For the FIML estimator $\tilde{\Pi}_0$, the result follows from equations <3.3.7.1>, <3.4.3.5>, and <3.4.3.8>: it is perhaps worthwhile to fill out the argument. By equation <3.4.3.5>,

$$n^{1/2}(\tilde{\theta} - \theta^0) \overset{a}{\approx} B(\alpha^0)n^{1/2}(\tilde{\alpha} - \alpha^0);$$

since B is block diagonal, and using equation <3.3.7.1>,

$$n^{1/2}(\tilde{\pi}_0 - \pi^0) \overset{a}{\approx} -(A^{0-1\prime} \otimes B^0)Hn^{1/2}(\tilde{\gamma} - \gamma^0)$$

$$\overset{a}{\approx} (A^{0-1\prime} \otimes B^0)H[H'(\Sigma^{0-1} \otimes B^{0\prime}M_XB^0)H]^{-1} \times$$

$$H'(A^{0-1}\Omega^{0-1} \otimes B^{0\prime})n^{-1/2}(I_m \otimes X')v_0.$$

<3.5.3.8>

Since both

$n^{1/2}\mathrm{vec}(\hat{\Pi} - \Pi^0)$ and $n^{1/2}\mathrm{vec}(\tilde{\Pi}_0 - \Pi^0)$

depend on

$n^{-1/2}(I_m \otimes X')v_0$,

it follows that they have a joint limiting normal distribution with mean zero and covariance matrix **V**, say.

Combining equations <3.5.3.5> and <3.5.3.8>, one obtains

$$n^{1/2}\mathrm{vec}(\hat{\Pi} - \tilde{\Pi}_0) \overset{a}{\approx} [I_{mk_1} - (A^{0-1\prime} \otimes B^0)H(H'(\Sigma^{0-1} \otimes B^{0\prime}M_XB^0)H)^{-1} \times$$

$$H'(A^{0-1}\Omega^{0-1} \otimes B^{0\prime}M_X)]n^{-1/2}(I_m \otimes M_X^{-1}X')v_0$$

$$= P(\alpha^0)n^{-\frac{1}{2}}(I_m \otimes M_X^{-1}X')v_0,$$

where $P(\alpha^0)$ is defined by equation <3.4.3.5>.

One can now return to equation <3.5.3.4> to deduce that

$$n^{\frac{1}{2}}(\gamma^\dagger - \gamma^0) \overset{a}{\approx} (H'(\Sigma^{0-1} \otimes \Pi^{0\prime}M_X\Pi^0)H)^{-1}\{n^{-\frac{1}{2}}H'(A^{0-1}\Omega^{0-1} \otimes \Pi^{0\prime}X')v_0$$
$$+ H'(\Sigma^{0-1}G_1' \otimes \Pi^{0\prime}M_X)P(\alpha^0)(I_m \otimes M_X^{-1})$$
$$\times\ n^{-\frac{1}{2}}(I_m \otimes X')v_0\},$$

which has a limit normal distribution with mean zero and covariance matrix **W**: it is easy to show that the cross-products between the two terms on the right hand side contribute nothing to the limiting covariance matrix, which can thus be written as

$$(H'(\Sigma^{0-1} \otimes \Pi^{0\prime}M_X\Pi^0)H)^{-1}$$
$$\times\ \{H'(\Sigma^{0-1} \otimes \Pi^{0\prime}M_X\Pi^0)H +$$
$$H'(\Sigma^{0-1}G_1' \otimes \Pi^{0\prime}M_X)\ P(\alpha^0)(\Omega^0 \otimes M_X^{-1})P'(\alpha^0)(G_1\Sigma^{0-1} \otimes M_X\Pi^0)H\}$$
$$\times\ (H'(\Sigma^{0-1} \otimes \Pi^{0\prime}M_X\Pi^0)H)^{-1}.$$

The first term in this expression is the covariance matrix

$\Psi(\tilde{\gamma};\gamma^0)$

of the limiting distribution of $\tilde{\gamma}$, and the second term is clearly positive semi-definite. Thus, $\gamma^\dagger$ is asymptotically inefficient relative to the FIML estimator $\tilde{\gamma}$.

One may conclude that Hendry's assertion that the asymptotic efficiency of estimators like $\gamma^\dagger$ defined in equation <3.5.2.1> depends on the efficiency of the reduced form parameter estimators is correct, but that this dependence is more complex than Hendry's assertion suggests.

3.5.4. Hendry's paper (Hendry [1976]) shows that if

$$\Sigma_0^* \xrightarrow{P} \Sigma^0, \quad \Pi_0^* \xrightarrow{P} \Pi^0,$$

then an estimator γ^{++} having the same asymmetric form as the FIML estimator $\tilde{\gamma}$

$$\gamma^{++} = -[H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'Z_1)H]^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'Z_1)h,$$

will have the same asymptotic properties as $\tilde{\gamma}$: given the initial estimators Σ_0^*, Π_0^*, this estimator γ^{++} can be obtained by an instrumental variables regression. There is a possible computational advantage in using only a "generalised least squares" regression, so that ways of correcting the symmetric estimator γ^{+} of equation <3.5.2.1> for the consequences of symmetry would be useful.

One simple possibility, considering equation <3.5.3.2>, is to add in a term involving

$$(V_0^* : 0) = Z_1 - X\Pi_0^*,$$

and redefine γ^{+} as

$$\begin{aligned}
\gamma^{+} &= -(H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H)^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')[(I_m \otimes Z_1)h \\
&\qquad + (I_m \otimes (V_0^*:0))H\gamma^*] \\
&= -(H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H)^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')[(I_m \otimes Z_1)h \\
&\qquad + (I_m \otimes Z_1)H\gamma^* - (I_m \otimes X\Pi_0^*)H\gamma^*] \\
&= -(H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H)^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')[(I_m \otimes Z_1)g_0^* \\
&\qquad - (I_m \otimes X\Pi_0^*)H\gamma^*] \\
&= \gamma^* - (H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H)^{-1}H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')u_0^*,
\end{aligned}$$

<3.5.4.1>

which is simply the two-step estimator of equation <3.5.1.5>; in this derivation, equations <3.5.1.3> and <3.5.1.4> have also been used.

The result shows that two-step estimation is perhaps a more direct way of getting estimators which are asymptotically equivalent to FIML estimators than by using symmetric FIML estimation; in addition, one can see how to obtain the two-step estimator $\hat{\hat{\gamma}}$ by a regression of

$$(I_m \otimes Z_1)h + (I_m \otimes (V_0^*:0))H\gamma^*$$

on

$$(I_m \otimes X\Pi_0^*)H$$

in the metric of $\Sigma_0^{*-1} \otimes I_n$.

There is also a trivial way of producing a symmetric FIML estimator: one simply recognises that an instrumental variables regression of the form

$$(W'X)^{-1}W'y$$

can be written as

$$(X'W(W'W)^{-1}W'X)^{-1}X'W(W'W)^{-1}W'y.$$

To apply this idea, define the matrices

$$\mathbf{S} = H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'Z_1)H,$$

$$\mathbf{R} = H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H$$

and the vector

$$\mathbf{d} = H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'Z_1)h.$$

Then the estimator

$$\gamma^{**} = -(\mathbf{S}'\mathbf{R}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{R}^{-1}\mathbf{d}$$

has exactly the same form as the FIML estimator $\tilde{\gamma}$, and will have the same limiting distribution.

Although this estimator γ^{**} can be produced by a regression, it is more complex than the estimator $\gamma^{\dagger}$ of equation <3.5.4.1>. The explicit two-step estimator is probably the most useful, since the convergent iterate $\hat{\hat{\gamma}}_\infty$ of the iteration scheme associated with this estimator satisfies the same equation as the FIML estimator $\tilde{\gamma}$. To display this, suppose that convergence occurs at iteration i, so that

$$\hat{\hat{\gamma}}_\infty = \hat{\hat{\gamma}}_{i+1} = \hat{\hat{\gamma}}_i ,$$

and that the structural form residual vector $\hat{u}_0$ associated with $\hat{\hat{\gamma}}_\infty$ satisfies

$$\hat{u}_0 = (I_m \otimes Z_1)\hat{g}_0 = (I_m \otimes Z_1)(H\hat{\hat{\gamma}}_\infty + h),$$

where

$$\hat{g}_0,\ \hat{\hat{\Sigma}}_0,\ \hat{\hat{\Omega}}_0$$

denote the estimators of g_0, Σ_0, Ω_0 associated with $\hat{\hat{\delta}}_\infty$. Then,

$$0 = \hat{\hat{\gamma}}_{i+1} - \hat{\hat{\gamma}}_i = -\ [H'(\hat{\hat{\Sigma}}_0^{-1} \otimes \hat{\hat{\Omega}}_0'X'X\hat{\hat{\Omega}}_0)H]^{-1}H'(\hat{\hat{\Sigma}}_0^{-1} \otimes \hat{\hat{\Omega}}_0'X')\hat{u}_0$$

$$= -[H'(\hat{\hat{\Sigma}}_0^{-1} \otimes \hat{\hat{\Omega}}_0'X'X\hat{\hat{\Omega}}_0)H]^{-1}H'(\hat{\hat{\Sigma}}_0^{-1} \otimes \hat{\hat{\Omega}}_0'X'Z_1)(H\hat{\hat{\gamma}}_\infty + h);$$

cancelling the inverse matrix in the square brackets produces an equation identical in form to that satisfied by the FIML estimator $\tilde{\gamma}$, equation <3.3.5.7>.

3.6. Minimum Chi-Squared Estimators

In this section, the minimum chi-squared estimators discussed in section 2.5. and the linearised minimum chi-squared estimators of section 2.7. are applied to the estimation of the parameters of the simultaneous equations model supposed to rule under the null hypothesis: this model was defined in subsection 3.1.3. as

$(I_m \otimes Z_1)g_o = u_o,$

$g_o = H\gamma + h.$

The main interest here lies in establishing the nature of these estimators in the simultaneous equations model, and how they relate to other estimators.

3.6.1. To obtain estimators of

$\alpha' = (\alpha_1' : \alpha_2') = (v'(\Omega_o) : \gamma')$

from the unrestricted maximum likelihood estimator $\hat{\theta}$ of

$\theta' = (\theta_1' : \theta_2') = (v'(\Omega_o) : \text{vec } \Pi_o'),$

the minimum chi-squared principle minimises

$(\hat{\theta} - \theta(\alpha))' I_n(\hat{\theta})(\hat{\theta} - \theta(\alpha))$

(see equation <2.5.1.1>). This expression collapses in the case of the simultaneous equations model, since $I_n(\hat{\theta})$ is block diagonal (see equation <3.4.2.3>), and trivially,

$\theta_1 = v(\Omega_o) = \alpha_1.$

The criterion function is thus

$$n^{-1}(\hat{\pi} - \theta_2(\gamma))'(\hat{\Omega}^{-1} \otimes X'X)(\hat{\pi} - \theta_2(\gamma)), \qquad \langle 3.6.1.1\rangle$$

where

$\hat{\pi} = \text{vec } \hat{\Pi} = \text{vec } ((X'X)^{-1}X'Y),$

the OLS estimator, and $\hat{\Omega}$ the corresponding unrestricted reduced form covariance matrix estimator,

$$\hat{\Omega} = n^{-1}Y'(I_n - X(X'X)^{-1}X')Y.$$

The function $\theta_2(\gamma)$ is an unpleasant chaining of

$$\pi_0 = \mathrm{vec}(-B_0A_0^{-1}),$$

a function of $g_0 = \mathrm{vec}\ C_0$, and

$$g_0 = H\gamma + h.$$

The first-order conditions for minimising <3.6.1.1> are

$$-2n^{-1}(D_\gamma\theta_2)'(\hat{\Omega} \otimes X'X)(\hat{\pi} - \theta_2(\gamma)) = 0,$$

which will require some kind of linearisation and iterative method for practical solution. It thus seems most sensible to turn directly to the consideration of the linearised minimum chi-squared estimator, which will in the current circumstance be found by minimising the analogue of

$$((\hat{\theta} - \theta^*) - B(\alpha^*)(\alpha^\nabla - \alpha^*))'I_n(\hat{\theta})((\hat{\theta} - \theta^*) - B(\alpha^*)(\alpha^\nabla - \alpha^*)),$$

which is equation <2.7.1.2>.

Neglecting the Ω_0-components of θ and α, and recalling from equation <3.4.3.8> that

$$D_\gamma\theta_2(\gamma) = -(A_0^{-1\prime} \otimes \Pi_0)H,$$

the appropriate minimand is

$$n^{-1}((\hat{\pi} - \pi_0^*) + (A_0^{*-1\prime} \otimes \Pi_0^*)H(\gamma^\nabla - \gamma^*))'(\hat{\Omega}^{-1} \otimes X'X) \times$$
$$((\hat{\pi} - \pi_0^*) + (A_0^{*-1\prime} \otimes \Pi_0^*)H(\gamma^\nabla - \gamma^*)),$$

π_0^* and γ^* being the appropriately "consistent and asymptotically normal" initial estimators of the null hypothesis model. It then follows that the linearised minimum chi-squared estimator of γ is

$$\gamma^{\triangledown} = \gamma^{*}+[H'(A_0^{*-1}\hat{\Omega}A_0^{*-1\prime} \otimes \Omega_0^{*\prime}X'X\Omega_0^{*})H]^{-1}H'(A_0^{*-1}\hat{\Omega} \otimes \Omega_0^{*\prime}X'X)(\hat{\pi} - \pi_0^{*}).$$

Such an estimator does not seem to have been suggested in the simultaneous equations literature; the "update" term can clearly be obtained by the regression of $(\hat{\pi} - \pi_0^{*})$ on

$$(A_0^{*-1\prime} \otimes X\Omega_0^{*})H$$

with respect to the metric $(\hat{\Omega} \otimes X'X)$.

<u>3.6.2.</u> Another way to approach minimum chi-squared estimation in the simultaneous equations model is to express the general null hypothesis model as in equation <2.1.1.2>:

$$\mathbf{H}_0: \quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha),$$

and then to note that the δ-component of β and the γ-component of α are linearly related under the null hypothesis by (equation <3.1.3.1>):

$$\delta = L\gamma + r.$$

The covariance matrix of the maximum likelihood estimator $\tilde{\delta}$ of δ from the alternative hypothesis model of equations <3.1.2.6> and <3.1.2.7>,

$$(I_m \otimes Z_1)g_1 = u_1,$$

$$g_1 = K\delta + k$$

is given by equation <3.4.3.10>:

$$\Psi(\tilde{\delta};\psi_1^0) = (K'(\Sigma^{0-1} \otimes \Omega^{0\prime}M_X\Omega^0)K)^{-1}.$$

The vector ψ_1^0 and its estimator are defined by equations <3.4.3.1> and <3.4.3.2>; an appropriate estimator of $\Psi(\tilde{\delta};\psi_1^0)$ is denoted

$$\Psi_n(\tilde{\delta};\tilde{\psi}_1) = n(K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)K)^{-1}.$$

By minimising

$$n^{-1}(\tilde{\delta} - L\gamma - r)'K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)K(\tilde{\delta} - L\gamma - r) \qquad \langle 3.6.2.1\rangle$$

another form of the linearised minimum chi-squared estimator is obtained,

$$\begin{aligned}\gamma^* &= (L'K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)KL)^{-1}L'K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)K(\tilde{\delta} - r) \\ &= (H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)H)^{-1}H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)K(\tilde{\delta} - r).\end{aligned}$$

It was observed in equations <3.1.4.1> and <3.1.4.2> that under the null hypothesis,

$$H = KL,$$

$$h = Kr + k,$$

so that

$$K(\tilde{\delta} - r) = K\tilde{\delta} + k - h = \tilde{g}_1 - h, \qquad \langle 3.6.2.2\rangle$$

and

$$(I_m \otimes \tilde{\Omega}_1)\tilde{g}_1 = \mathrm{vec}(\tilde{\Pi}_1\tilde{A}_1 + \tilde{B}_1) = 0.$$

Thus, the expression for γ^* collapses to

$$\gamma^* = -(H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)H)^{-1}H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1X'X\tilde{\Omega}_1)h, \qquad \langle 3.6.2.3\rangle$$

which is very similar in form to the symmetric FIML estimators discussed in subsections 3.5.2. and 3.5.4.; this estimator, however, clearly has from the general results the same limiting distribution as the maximum likelihood estimator $\tilde{\gamma}$. The estimator may be obtained by means of a regression of "fitted values" on "fitted values" - that is, of

$$-(I_m \otimes X\tilde{\Omega}_1)h$$

on

$$(I_m \otimes X\tilde{\Omega}_1)H$$

in the metric of $(\tilde{\Sigma}_1^{-1} \otimes I_n)$.

<u>3.6.3.</u> It is interesting to investigate what happens to this linearised minimum chi-squared estimator when the alternative

hypothesis model is just identified, that is, when the number of columns of K, q_1, equals mk_1, the number of elements of π_1. In this case,

$$(I_m \otimes \Omega_1)K$$

is square and non-singular, so that the FIML estimator of δ can be obtained from equation <3.3.4.7> as

$$\tilde{\delta} = -\,[K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'Z_1)K]^{-1}K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'Z_1)k$$

$$= -\,[(I_m \otimes X'Z_1)K]^{-1}(I_m \otimes X'Z_1)k, \qquad \langle 3.6.3.1\rangle$$

the well known indirect least squares estimator, which can also be obtained from solving the equations

$$\hat{\pi} = \phi_2(\tilde{\delta}).$$

This reveals that $\tilde{\Omega}_1$ coincides with the least squares estimator $\hat{\Omega}$: let

$$P_X = X(X'X)^{-1}X';$$

then

$$X\tilde{\Omega}_1 \equiv X\hat{\Omega} = P_X Z_1,$$

so that one can write

$$\gamma^* = -[H'(\tilde{\Sigma}_1^{-1} \otimes Z_1'P_XZ_1)H]^{-1}H'(\tilde{\Sigma}_1^{-1} \otimes Z_1'P_XZ_1)h. \qquad \langle 3.6.3.2\rangle$$

This gives an explicit (and simple) algebraic form for the "constrained indirect least squares" estimator introduced by Wegge [1978]. If one then chooses $\tilde{\Sigma}_1$ to be the two-stage least squares estimator, γ^* becomes the well known three-stage least squares estimator.

3.7. The Limited Information Maximum Likelihood Estimator.

3.7.1. The estimation of a simultaneous equations model in which m^*, say, equations are overidentified, and the remaining $m - m^*$ are just-identified provides a useful link between the FIML estimator of the whole system of equations, and single-equation or limited information estimators such as limited information maximum likelihood (LIML) and two-stage least squares (see Hendry [1976], Court [1973,1974]), for when there is only one overidentified equation in the model, such estimators are asymptotically equivalent to FIML. Another way of making the same point is to observe that limited information estimators assume that the simultaneous equations model has this "one over-identified equation" structure; if, in fact, the model does not have this structure, then limited information estimators are inefficient relative to the FIML estimator.

It is of some interest to see how far one can go with general constraint parameter restrictions in LIML estimation, thus providing a mild generalisation of the usual case of exclusion restrictions and a single unit normalisation rule. In addition, inference on the parameters of a single over-identified equation then becomes possible, in a manner consistent with inference for the complete system of equations: this will be discussed in Chapter 5. Finally, the tests of identification discussed in the econometric literature have been based on the use of the LIML estimator, so that the results of this section will provide a basis for the next Chapter.

The analysis of the LIML estimator which follows is similar to that given by Hendry [1976, section 3.1].

<u>3.7.2.</u> It will be convenient to work with the null hypothesis model as described by equations <3.1.3.2>, <3.1.3.3>, <3.1.3.5> and <3.1.3.6>: the reduced form is

$Y = X\Pi_0 + V_0,$

the structural form

$YA_0 + XB_0 = U_0$

or

$Z_1C_0 = U_0,$

$(I_m \otimes Z_1)g_0 = u_0$

with

$g_0 = H\gamma + h.$

This model now needs partitioning into the first and remaining m − 1 equations: so, let

$A_0 = (a_{0.1} : A_{02}), \quad a'_{0.1} = (a_{0.11} : a'_{0.12}),$

where $a_{0.12}$ is $(m - 1) \times 1$;

$C_0 = (c_{0.1} : C_{02}), \quad \Pi_0 = (\pi_{0.1} : \Pi_{02}),$

$g'_0 = (c'_{0.1} : (\text{vec } C_{02})') = (c'_{0.1} : g'_{02}),$

$Y = (y_1 : Y_2), \quad V_0 = (v_{0.1} : V_{02}), \quad U_0 = (u_{0.1} : U_{02}),$

and

$$\text{vec } V_o = v_o = \begin{bmatrix} v_{o.1} \\ v_{o2} \end{bmatrix}, \quad \text{vec } U_o = u_o = \begin{bmatrix} u_{o.1} \\ u_{o2} \end{bmatrix}$$

so that

$$v_{o2} = \text{vec } V_{o2}, \quad u_{o2} = \text{vec } U_{o2}.$$

It will be convenient to allow the $m - 1$ just-identified equations to be expressed in reduced form:

$$Y_2 = X\Pi_{o2} + V_{o2}, \qquad \text{<3.7.2.1>}$$

whilst the single over-identified equation is

$$Z_1 c_{o.1} = u_{o.1} \cdot \qquad \text{<3.7.2.2>}$$

The "free" parameters of the first equation may be denoted by $\gamma_{.1}$, of dimension $q_{o1} \times 1$, so that

$$\gamma' = (\gamma'_{.1} : (\text{vec } \Pi_{o2})') = (\gamma'_{.1} : \pi'_{o2}).$$

The matrix H of the restrictions

$$g_o = H\gamma + h$$

is then block-diagonal:

$$H = \bigoplus_{i=1}^{2} H_{ii},$$

with H_{11} being $(m + k_1) \times q_{o1}$, and H_{22} $(m - 1)(m + k_1) \times (m - 1)k_1$;

h is similarly partitioned as

$$h' = (h'_{.1} : h'_{.2}),$$

with $h_{.1}$ being $(m + k_1) \times 1$ and $h_{.2}$ $(m - 1)(m + k_1) \times 1$.

Thus, the restrictions on the first equation may be expressed as

$$c_{o.1} = H_{11}\gamma_{.1} + h_{.1}, \qquad \text{<3.7.2.3>}$$

whilst the structure of equation <3.7.2.1> shows that

$$C'_{o2} = (O_{m-1,1} : I_{m-1} : -\Pi'_{o2}),$$

so that

$$C_{o2} = \begin{bmatrix} O_{mk_1} \\ -I_{k_1} \end{bmatrix} \Pi_{o2} + \begin{bmatrix} O_{1,m-1} \\ I_{m-1} \\ O_{k_1,m-1} \end{bmatrix}.$$

Then, one can write

$$g_{o2} = \left(I_{m-1} \otimes \begin{bmatrix} O_{mk_1} \\ -I_{k_1} \end{bmatrix}\right) \text{vec}\ \Pi_{o2} + \text{vec} \begin{bmatrix} O_{1,m-1} \\ I_{m-1} \\ O_{k_1,m-1} \end{bmatrix} \qquad \langle 3.7.2.4\rangle$$

which establishes the structure of H_{22} and $h_{.2}$ by comparing this expression with

$$g_{o2} = H_{22}\pi_{o2} + h_{.2}, \qquad \langle 3.7.2.5\rangle$$

the appropriate sub-equations of

$$g_o = H\gamma + h.$$

There are some useful consequences of this discussion: the matrix H_{22} satisfies

$$(I_{m-1} \otimes \Omega_o)H_{22} = \left[I_{m-1} \otimes (\Pi_o : I_{k_1}) \begin{bmatrix} O_{mk_1} \\ -I_{k_1} \end{bmatrix}\right] = -(I_{m-1} \otimes I_{k_1}),$$

$$A'_{o2} = (O_{m-1,1} : I_{m-1}),$$

and

$$(I_{m-1} \otimes Z_1)h_{.2} = \text{vec}\left(Z_1 \begin{bmatrix} O_{1,m-1} \\ I_{m-1} \\ O_{k_1,m-1} \end{bmatrix}\right) = \text{vec}\ Y_2. \qquad \langle 3.7.2.6\rangle$$

<u>3.7.3.</u> To specialise the FIML estimators $\tilde{\Omega}_o$, $\tilde{\Pi}_o$ and $\tilde{\gamma}$ of the null hypothesis model, it is helpful to summarise the first-order conditions for maximising

$n^{-1}l_n(y;\theta)$ subject to $\pi_0 = \theta_2(\gamma)$,

which may be deduced from equations <3.3.5.1>-<3.3.5.6>, allowing for the use of the "short" form $\theta = \theta(\alpha)$ of the restrictions of the null hypothesis model. The Lagrange multiplier of this problem is denoted τ_2.

$$0 = -\tfrac{1}{2}\mathbf{D}_m(\tilde{\Omega}_0^{-1} \otimes \tilde{\Omega}_0^{-1})\mathbf{D}_m' v[\tilde{\Omega}_0 - n^{-1}(Y-X\tilde{\Pi}_0)'(Y-X\tilde{\Pi}_0)] \qquad \text{<3.7.3.1>}$$

$$0 = n^{-1}(\tilde{\Omega}_0^{-1} \otimes X')\text{vec}(Y - X\tilde{\Pi}_0) + \tilde{\tau}_2 \qquad \text{<3.7.3.2>}$$

$$0 = H'(\tilde{A}_0^{-1} \otimes \tilde{B}_0')\tilde{\tau}_2 \qquad \text{<3.7.3.3>}$$

$$0 = \tilde{\pi}_0 - \theta_2(\tilde{\gamma}). \qquad \text{<3.7.3.4>}$$

Partitioning up $\tilde{\tau}_2$ to match that of γ,

$$\tilde{\tau}_2' = (\tilde{\tau}_{2.1}' : \tilde{\tau}_{2.2}'),$$

one can write

$$H'(\tilde{A}_0^{-1} \otimes \tilde{B}_0')\tilde{\tau}_2 = \begin{bmatrix} H_{11}' : 0 \\ 0 : H_{22}' \end{bmatrix} \begin{bmatrix} \tilde{a}_{0.11}^{-1}\tilde{B}_0' & : & 0 \\ -\tilde{a}_{0.11}^{-1}\tilde{a}_{0.12}' \otimes \tilde{B}_0' : I_{m-1} & \otimes & \tilde{B}_0' \end{bmatrix} \begin{bmatrix} \tilde{\tau}_{2.1} \\ \tilde{\tau}_{2.2} \end{bmatrix} = 0,$$

and establish that

$$\tilde{\tau}_{2.2} = (\tilde{a}_{0.11}^{-1}\tilde{a}_{0.12} \otimes I_{k_1})\tilde{\tau}_{2.1}. \qquad \text{<3.7.3.5>}$$

This result will be needed later.

To obtain the LIML estimators of $\gamma_{.1}$ and π_{02}, one can simply partition (according to $\gamma_{.1}$ and π_{02}) the FIML "normal equations" which stem from equations <3.7.3.2> and <3.7.3.3>: the normal equations are given by equation <3.3.5.7> as

$$[H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)H]\tilde{\gamma} = -H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{B}_0'X'Z_1)h.$$

The partitioning will require the partitioning of Σ_0 by the first and remaining $(m - 1)$ rows and columns:

$$\tilde{\Sigma}_0 = \begin{bmatrix} \tilde{\sigma}_{0.11} : \tilde{\sigma}_{0.1}' \\ \tilde{\sigma}_{0.1} : \tilde{\Sigma}_{0.2} \end{bmatrix}, \qquad \tilde{\Sigma}_0^{-1} = \begin{bmatrix} \tilde{\sigma}_0^{.11} : \tilde{\sigma}_0^{.1\prime} \\ \tilde{\sigma}_0^{.1} : \tilde{\Sigma}_0^{.2} \end{bmatrix}. \qquad \text{<3.7.3.6>}$$

The partitioning yields, using equation <3.7.2.6>,

$$\begin{bmatrix} \tilde{\sigma}_0^{\cdot 11}H_{11}'\tilde{\Omega}_0'X'Z_1H_{11} & : & -H_{11}'(\tilde{\sigma}_0^{\cdot 1\prime} \otimes \tilde{\Omega}_0'X'X) \\ -(\tilde{\sigma}_0^{\cdot 1} \otimes X'Z_1)H_{11} & : & (\tilde{\Sigma}_0^{\cdot 2} \otimes X'X) \end{bmatrix}\begin{bmatrix} \tilde{\gamma}_{.1} \\ \tilde{\pi}_{02} \end{bmatrix}$$

$$= -\begin{bmatrix} \tilde{\sigma}_0^{\cdot 11}H_{11}'\tilde{\Omega}_0'X'Z_1h_{.1} + H_{11}'(\tilde{\sigma}_0^{\cdot 1\prime} \otimes \tilde{\Omega}_0'X')\mathrm{vec}\ Y_2 \\ -(\tilde{\sigma}_0^{\cdot 1} \otimes X'Z_1)h_{.1} - (\tilde{\Sigma}_0^{\cdot 2} \otimes X')\mathrm{vec}\ Y_2 \end{bmatrix}. \quad \text{<3.7.3.7>}$$

Solving the second equation for $\tilde{\pi}_{02}$ produces

$$\tilde{\pi}_{02} = \hat{\pi}_2 + (\tilde{\Sigma}_0^{\cdot 2} \otimes X'X)^{-1}(\tilde{\sigma}_0^{\cdot 1} \otimes X'Z_1)(H_{11}\tilde{\gamma}_{.1} + h_{.1})$$

$$= \hat{\pi}_2 - (\tilde{\sigma}_{0.11}^{-1}\tilde{\sigma}_{0.1} \otimes (X'X)^{-1}X'Z_1)\tilde{c}_{0.1} \quad \text{<3.7.3.8>}$$

on exploiting the partitioned inverse relationship

$$\tilde{\sigma}_0^{\cdot 1}\tilde{\sigma}_{0.11} + \tilde{\Sigma}_0^{\cdot 2}\tilde{\sigma}_{0.1} = 0; \quad \text{<3.7.3.9>}$$

$\hat{\pi}_2$ is the vec of the OLS estimator

$$\hat{\Pi}_2 = (X'X)^{-1}X'Y_2,$$

a submatrix of

$$\hat{\Pi} = (X'X)^{-1}X'Y.$$

Substituting the solution $\tilde{\pi}_{02}$ of equation <3.7.3.8> into the first equation of <3.7.3.7> yields, after exploiting the partitioned inverse result <3.7.3.9> again,

$$\tilde{\sigma}_{0.11}^{-1}H_{11}'\tilde{\Omega}_0'X'Z_1H_{11}\tilde{\gamma}_{.1} = -\tilde{\sigma}_{0.11}^{-1}H_{11}'\tilde{\Omega}_0'X'Z_1h_{.1},$$

or, deleting $\tilde{\sigma}_{0.11}^{-1}$ and using equation <3.7.2.3>,

$$H_{11}'\tilde{\Omega}_0'X'Z_1(H_{11}\tilde{\gamma}_{.1} + h_{.1}) = 0$$

or

$$H_{11}'\tilde{\Omega}_0'X'Z_1\tilde{c}_{0.1} = 0. \quad \text{<3.7.3.10>}$$

The equation <3.7.3.8> above can be seen to be equivalent to

$$\tilde{\Pi}_{02} = \hat{\Pi}_2 - \tilde{\sigma}_{0.11}^{-1}(X'X)^{-1}X'Z_1\tilde{c}_{0.1}\tilde{\sigma}_{0.1}' \quad \text{<3.7.3.11>}$$

$$= \hat{\Pi}_2 - \hat{\Omega}\tilde{c}_{0.1}\tilde{\sigma}_{0.1}'\tilde{\sigma}_{0.11}^{-1},$$

where

$$\hat{Q} = (\hat{\Pi} : I_{k_1}).$$

Recognising that

$$\tilde{\Sigma}_0 = n^{-1}\tilde{A}_0'\tilde{\Omega}_0\tilde{A}_0 = n^{-1}\tilde{U}_0'\tilde{U}_0 = n^{-1}\tilde{C}_0'Z_1'Z_1\tilde{C}_0$$

$$= n^{-1}\begin{bmatrix} \tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1} & : & \tilde{c}_{0.1}'Z_1'Z_1\tilde{C}_{02} \\ \tilde{C}_{02}'Z_1'Z_1\tilde{c}_{0.1} & : & \tilde{C}_{02}'Z_1'Z_1\tilde{C}_{02} \end{bmatrix},$$

and that

$$Z_1\tilde{C}_{02} = Y_2 - X\tilde{\Pi}_{02},$$

one then obtains

$$\tilde{\Pi}_{02} = \hat{\Pi}_2 - (\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1})^{-1}\hat{Q}\tilde{c}_{0.1}\tilde{c}_{0.1}'Z_1'(Y_2 - X\tilde{\Pi}_{02}).$$

Transposing and multiplying through by $X'Z_1\tilde{c}_{0.1}$, one obtains

$$\tilde{\Pi}_{02}'X'Z_1\tilde{c}_{0.1} = \hat{\Pi}_2'X'Z_1\tilde{c}_{0.1}$$

$$- (\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1})^{-1}(Y_2-X\tilde{\Pi}_{02})'Z_1\tilde{c}_{0.1}\tilde{c}_{0.1}'\hat{Q}'X'Z_1\tilde{c}_{0.1}$$

$$= \hat{\Pi}_2'X'Z_1\tilde{c}_{0.1} - \nu Y_2'Z_1\tilde{c}_{0.1} + \nu\tilde{\Pi}_{02}'X'Z_1\tilde{c}_{0.1},$$

on defining

$$\nu = (\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1})^{-1}\tilde{c}_{0.1}'\hat{Q}'X'Z_1\tilde{c}_{0.1}$$

$$= (\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1})^{-1}\tilde{c}_{0.1}'Z_1'P_XZ_1\tilde{c}_{0.1}, \qquad \langle 3.7.3.12\rangle$$

where

$$P_X = X(X'X)^{-1}X'.$$

Hence,

$$(1 - \nu)\tilde{\Pi}_{02}'X'Z_1\tilde{c}_{0.1} = Y_2'(P_X - \nu I_n)Z_1\tilde{c}_{0.1}, \qquad \langle 3.7.3.13\rangle$$

since

$$X\hat{\Pi}_2 = P_XY_2.$$

At this point, it seems essential to declare explicitly the normalisation rule for the first equation: $a_{0.11}$ is a known positive number, taken to be unity without loss of generality,

$a_{0.11} = 1.$

This ensures that the first row of H_{11} is null, and that the first element of $h_{.1}$ is unity:

$H'_{11} = (0_{1,m+k_1} : H'_{11.2}).$

Then,

$$H'_{11}\tilde{\Omega}'_0 = H'_{11}\begin{bmatrix}\tilde{\Pi}'_0 \\ I_{k_1}\end{bmatrix} = (0_{1,m+k_1} : H'_{11.2})\begin{bmatrix}\tilde{\pi}'_{0.1} \\ \tilde{\Pi}'_{02} \\ I_{k_1}\end{bmatrix} = H'_{11.2}\begin{bmatrix}\tilde{\Pi}'_{02} \\ I_{k_1}\end{bmatrix}$$

and so, by equation <3.7.3.10>,

$$0 = H'_{11}\tilde{\Omega}'_0 X' Z_1 \tilde{c}_{0.1} = H'_{11.2}\begin{bmatrix}\tilde{\Pi}'_{02} \\ I_{k_1}\end{bmatrix} X' Z_1 \tilde{c}_{0.1}. \qquad \text{<3.7.3.14>}$$

Multiplying through by $(1 - \nu)$, and noting that

$$(1 - \nu)X' = X'(P_X - \nu I_n), \qquad \text{<3.7.3.15>}$$

one obtains from equations <3.7.3.13> and <3.7.3.14>,

$$\begin{aligned} 0 &= H'_{11.2}\begin{bmatrix}(1 - \nu)\tilde{\Pi}'_{02} \\ (1 - \nu) I_{k_1}\end{bmatrix} X' Z_1 \tilde{c}_{0.1} \\ &= H'_{11.2}\begin{bmatrix} Y'_2 \\ X' \end{bmatrix}(P_X - \nu I_n) Z_1 \tilde{c}_{0.1} \\ &= H'_{11} Z'_1 (P_X - \nu I_n) Z_1 \tilde{c}_{0.1}. \end{aligned} \qquad \text{<3.7.3.16>}$$

Let

$$H^*_{11} = (h_{.1} : H_{11}),\ \gamma^{*\prime}_{.1} = (1 : \gamma'_{.1}) : \qquad \text{<3.7.3.17>}$$

then

$$\tilde{c}_{0.1} = H_{11}\gamma_{.1} + h_{.1} = H^*_{11}\gamma^*_{.1}.$$

Since

$$0 = \tilde{\Omega}_0 \tilde{c}_{0.1} = \tilde{\Omega}_0 h_{.1} + \tilde{\Omega}_0 H_{11} \tilde{\gamma}_{.1},$$

the columns of $\tilde{\Omega}_0 H^*_{11}$ are linearly dependent on those of

$\tilde{\Omega}_0 H_{11}$: equations <3.7.3.14> and <3.7.3.16> above then imply that

$$0 = H_{11}^{*\prime} Z_1'(P_X - \nu I_n) Z_1 \tilde{c}_{0.1}$$
$$= H_{11}^{*\prime} Z_1'(P_X - \nu I_n) Z_1 H_{11}^{*} \gamma^{*}_{.1}, \qquad \text{<3.7.3.18>}$$

showing that $\tilde{\gamma}^{*}_{.1}$ is found as a (suitably normalised) characteristic vector corresponding to some characteristic root ν of the generalised characteristic value problem

$$H_{11}^{*\prime} Z_1' P_X Z_1 H_{11}^{*} \gamma^{*}_{.1} = \nu H_{11}^{*\prime} Z_1' Z_1 H_{11}^{*} \gamma^{*}_{.1}. \qquad \text{<3.7.3.19>}$$

To establish which particular characteristic root is required to maximise the log-likelihood function, it is necessary, as a preliminary, to obtain an expression for the Lagrange multiplier $\tilde{\tau}_{2.1}$: this is also necessary for the construction of single equation Lagrange Multiplier tests of overidentifying restrictions.

3.7.4. From equation <3.7.3.2>, the Lagrange multiplier of the problem, $\tilde{\tau}_2$, is given by

$$\tilde{\tau}_2 = -n^{-1}(\tilde{\Omega}_0^{-1} \otimes X')\tilde{v}_0, \qquad \text{<3.7.4.1>}$$

where $\tilde{v}_0$ is the reduced form residual vector associated with $\tilde{\Pi}_0$:

$$\tilde{v}_0 = \text{vec } \tilde{V}_0 = \text{vec}(Y - X\tilde{\Pi}_0).$$

The Lagrange multiplier $\tilde{\tau}_2$ is partitioned into a component, $\tilde{\tau}_{2.1}$, relating to the restrictions on the first equation, and the remainder, $\tilde{\tau}_{2.2}$, relating to the $m - 1$ just-identified equations; these subvectors satisfy equation <3.7.3.5>, but since it has been assumed that

$$a_{0.11} = 1,$$

as a normalisation rule, this relationship is written as

$$\tilde{\tau}_{2.2} = (\tilde{a}_{0.12} \otimes I_{k_1})\tilde{\tau}_2. \qquad \langle 3.7.4.2\rangle$$

Noting that

$\Omega_0^{-1} = A_0'\Sigma_0^{-1}A_0, \quad v_0 = (A_0^{-1\prime} \otimes I_n)u_0,$

and, since $\tilde{\tau}_{2.1}$ is $k_1 \times 1$, $\tilde{\tau}_2$ is $mk_1 \times 1$,

$\tilde{\tau}_{2.1} = (I_{k_1} : 0_{k_1,(m-1)k_1})\tilde{\tau}_2 = (e_1' \otimes I_{k_1})\tilde{\tau}_2,$

where e_1 is the first m-dimensional coordinate vector, one can write, using equation <3.7.4.1> above,

$\tilde{\tau}_{2.1} = -n^{-1}(e_1'\tilde{A}_0' \otimes I_{k_1})(\tilde{\Sigma}_0^{-1} \otimes X')\tilde{u}_0.$

Exploiting the structure of $\tilde{\Sigma}_0^{-1}$ given by equation <3.7.3.6> and

$$A_0 = \begin{bmatrix} 1 & : & 0 \\ a_{0.12} & : & I_{m-1} \end{bmatrix}$$

gives

$e_1'\tilde{A}_0'\tilde{\Sigma}_0^{-1} = (\tilde{\sigma}_0^{.11} : \tilde{\sigma}_0^{.1\prime}),$

so that

$$\begin{aligned}\tilde{\tau}_{2.1} &= -n^{-1}((\tilde{\sigma}_0^{.11} : \tilde{\sigma}_0^{.1\prime}) \otimes X')\tilde{u}_0 \\ &= -n^{-1}(\tilde{\sigma}_0^{.11}X'\tilde{u}_{0.1} + (\tilde{\sigma}_0^{.1\prime} \otimes X')\tilde{u}_{02}),\end{aligned}$$

using the decomposition of u_0 given in subsection 3.7.2. .

To proceed any further, an expression for the residual vector $\tilde{u}_{02}$ is needed: because the just-identified equations are in reduced form,

$$\begin{aligned}\tilde{u}_{02} &= y_2 - (I_{m-1} \otimes X)\tilde{\pi}_{02} \\ &= y_2 - (I_{m-1} \otimes X)[\hat{\pi}_2 - (\tilde{\sigma}_{0.11}^{-1}\tilde{\sigma}_{0.1} \otimes (X'X)^{-1}X'Z_1)\tilde{c}_{0.1}],\end{aligned}$$

using equation <3.7.3.11>, and

$\hat{u}_{02} = y_2 - (I_{m-1} \otimes X)\hat{\pi}_2,$

so that

$$\tilde{u}_{02} = \hat{u}_{02} + (\tilde{\sigma}_{0.11}^{-1}\tilde{\sigma}_{0.1} \otimes P_X Z_1)\tilde{c}_{0.1},$$

where

$$P_X = X(X'X)^{-1}X'.$$

The residual vector $\hat{u}_{02}$ and $(I_{m-1} \otimes X')$ are mutually orthogonal, so that

$$(\tilde{\sigma}_{0.1}' \otimes X')\tilde{u}_{02} = \tilde{\sigma}_{0.11}^{-1}\tilde{\sigma}_0^{.1\prime}\tilde{\sigma}_{0.1}X'Z_1\tilde{c}_{0.1},$$

and hence

$$\begin{aligned}\tilde{\tau}_{2.1} &= -n^{-1}(\tilde{\sigma}_0^{.11} + \tilde{\sigma}_{0.11}^{-1}\tilde{\sigma}_0^{.1\prime}\tilde{\sigma}_{0.1})X'Z_1\tilde{c}_{0.1}\\ &= -n^{-1}\tilde{\sigma}_{0.11}^{-1}X'Z_1\tilde{c}_{0.1}\\ &= -n^{-1}\tilde{\sigma}_{0.11}^{-1}X'\tilde{u}_{0.1} \qquad \text{<3.7.4.3>}\end{aligned}$$

using the partitioned inverse result

$$\tilde{\sigma}_0^{.11}\tilde{\sigma}_{0.11} + \tilde{\sigma}_0^{.1\prime}\tilde{\sigma}_{0.1} = 1,$$

and the fact that

$$Z_1\tilde{c}_{0.1} = \tilde{u}_{0.1}.$$

3.7.5. The log-likelihood function for this problem is that given by equation <3.3.6.1>:

$$n^{-1}l_n(y;\theta) = -ms - \tfrac{1}{2}\log\det\Omega_0 - \tfrac{1}{2}n^{-1}\mathrm{tr}[\Omega_0^{-1}(Y - X\Pi_0)'(Y - X\Pi_0)] \qquad \text{<3.7.5.1>}$$

whose value at the maximum likelihood estimators $\tilde{\Omega}_0$, $\tilde{\Pi}_0$ is

$$n^{-1}l_n(y;\tilde{\theta}) = -m(s+\tfrac{1}{2}n^{-1}) - \tfrac{1}{2}\log\det\tilde{\Omega}_0, \qquad \text{<3.7.5.2>}$$

where

$$\tilde{\Omega}_0 = n^{-1}\tilde{V}_0'\tilde{V}_0.$$

By rewriting the expression for $n^{-1}l_n(y;\tilde{\theta})$ in terms of the characteristic root ν of equation <3.7.3.19>, it will become apparent which choice of ν maximises the log-likelihood function.

The argument is based on the fact that the Lagrange multiplier τ_2 can be regarded as the vec of a $k_1 \times m$ matrix $\mathbf{T}$, with

$$\mathbf{T} = (\tau_{2.1} : \mathbf{T}_2),$$

so that

$$\tau_{2.2} = \text{vec } \mathbf{T}_2.$$

The relationship <3.7.4.2> between $\tilde{\tau}_{2.1}$ and $\tilde{\tau}_{2.2}$ can then be expressed as

$$\tilde{\mathbf{T}}_2 = \tilde{\tau}_{2.1}\tilde{a}'_{0.1},$$

giving

$$\mathbf{T} = \tilde{\tau}_{2.1}a',$$

where a' temporarily denotes

$$a' = (1 : \tilde{a}'_{0.1}),$$

and the equation <3.7.3.2> of the first-order conditions can be expressed as

$$n^{-1}X'(Y - X\tilde{\Pi}_0)\tilde{\Omega}_0^{-1} + \tilde{\mathbf{T}} = 0.$$

The solution of this equation for $\tilde{\Pi}_0$ is

$$\begin{aligned}\tilde{\Pi}_0 &= (X'X)^{-1}X'Y + n(X'X)^{-1}\tilde{\mathbf{T}}\tilde{\Omega}_0 \\ &= \hat{\Pi} + n(X'X)^{-1}\tilde{\mathbf{T}}\tilde{\Omega}_0;\end{aligned}$$

noting that

$$\tilde{V}_0 = Y - X\tilde{\Pi}_0,$$

$$\hat{V} = Y - X\hat{\Pi},$$

one obtains

$$\begin{aligned}\tilde{V}_0 &= \hat{V} - nX(X'X)^{-1}\tilde{\mathbf{T}}\tilde{\Omega}_0 \\ &= \hat{V} - X(X'X)^{-1}\tilde{\mathbf{T}}\tilde{V}_0'\tilde{V}_0,\end{aligned}$$

so that

$$\tilde{V}_0'\tilde{V}_0 = \hat{V}'\hat{V} + \tilde{V}_0'\tilde{V}_0\tilde{\mathbf{T}}'(X'X)^{-1}\tilde{\mathbf{T}}\tilde{V}_0'\tilde{V}_0.$$

Using equation <3.7.4.3>, $\tilde{\Upsilon}$ can be expressed as

$$\tilde{\Upsilon} = -n^{-1}\tilde{\sigma}_{0.11}^{-1}X'Z_1\tilde{c}_{0.1}a',$$

so that

$$\hat{V}'\hat{V} = \tilde{V}_0'\tilde{V}_0 - n^{-2}\tilde{\sigma}_{0.11}^{-2}\tilde{V}_0'\tilde{V}_0 a\tilde{c}_{0.1}'Z_1'X(X'X)^{-1}X'Z_1\tilde{c}_{0.1}a'\tilde{V}_0'\tilde{V}_0.$$

Before taking determinants, note that

$$\tilde{\sigma}_{0.11} = n^{-1}\tilde{u}_{0.1}'\tilde{u}_{0.1} = n^{-1}\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1}$$

and that, from equation <3.7.3.12>,

$$\nu = (\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1})^{-1}\tilde{c}_{0.1}'Z_1'X(X'X)^{-1}X'Z_1\tilde{c}_{0.1}.$$

Let

$$\mathbf{b} = (\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1})^{-1}\nu: \qquad \text{<3.7.5.3>}$$

then,

$$\hat{V}'\hat{V} = \tilde{V}_0'\tilde{V}_0[I_m - \mathbf{b}\mathbf{a}\mathbf{a}'\tilde{V}_0'\tilde{V}_0];$$

to find the determinant of this expression, one can use the result that for arbitrary m-vectors **c**, **d**,

$$\det(\mathbf{I}_m + \mathbf{c}\mathbf{d}') = 1 + \mathbf{c}'\mathbf{d}:$$

see for example, Rao [1973,p32].

Thus,

$$\det \hat{V}'\hat{V} = \det \tilde{V}_0'\tilde{V}_0(1 - \mathbf{b}\mathbf{a}'\tilde{V}_0'\tilde{V}_0\mathbf{a}):$$

in this expression, $\tilde{V}_0\mathbf{a}$ is actually the matrix $\tilde{V}_0$ multiplied by the first column of the matrix $\tilde{A}_0$; hence,

$$\tilde{V}_0\mathbf{a} = \tilde{u}_{0.1} = Z_1\tilde{c}_{0.1},$$

and then, using the definition <3.7.5.3>,

$$\mathbf{b}\tilde{c}_{0.1}'Z_1'Z_1\tilde{c}_{0.1} = \nu.$$

Overall, then,

$$\det \tilde{V}_0'\tilde{V}_0 = (1 - \nu)^{-1}\det \hat{V}'\hat{V}.$$

Since the latter determinant is a function of the observations only, the log-likelihood function of equation <3.7.5.2> equals

$$n^{-1}l_n(y;\tilde{\theta}) = -m(s+\tfrac{1}{2}n^{-1}) - \tfrac{1}{2}\log\det n^{-1}\tilde{V}_0'\tilde{V}_0$$

$$= -m(s+\tfrac{1}{2}n^{-1}) - \tfrac{1}{2}\log\det n^{-1}\hat{V}'\hat{V} + \tfrac{1}{2}\log(1 - \nu)$$

which is clearly maximised by choosing the smallest root

$$\tilde{\nu} = \nu_{min}$$

of the determinental equation arising from equation <3.7.3.18>:

$$\det H_{11}^{*\prime}Z_1'(P_X - \nu I_n)Z_1H_{11}^* = 0,$$

and choosing $\tilde{\gamma}^*_{.1}$ (and hence $\tilde{\gamma}_{.1}$) to be the corresponding, suitably normalised, characteristic vector. Thus,

$$n^{-1}l_n(y;\tilde{\theta}) = -m(s+\tfrac{1}{2}n^{-1}) - \tfrac{1}{2}\log\det n^{-1}\hat{V}'\hat{V} + \tfrac{1}{2}\log(1 - \tilde{\nu}).$$

<3.7.5.4>

<u>3.7.6.</u> The estimator derived above is a straightforward generalisation of the usual LIML estimator, and will collapse to this when the matrix H_{11}^* is block diagonal with respect to the endogenous and exogenous variables. One can therefore use the LIML estimator in quite general circumstances, with comparatively little change from the usual arguments, although the derivation given above is somewhat involved, as seems to be usual with LIML estimation. However, a link with the traditional approach of Anderson and Rubin [1949] can be obtained, and which yields a shorter derivation of the LIML estimator $\tilde{\gamma}_{.1}$ and the associated Lagrange multiplier $\tilde{\tau}_{2.1}$.

The Anderson and Rubin approach requires the maximisation of

$n^{-1} l_n(y;\theta)$

given by equation <3.7.5.1> subject only to the restrictions imposed on the first equation by

$c_{0.1} = H_{11}\gamma_{.1} + h_{.1}$;

in contrast to the preceding arguments in this section, the structural parameter restrictions are imposed via the first column of the relation

$\Pi_0 A_0 + B_0 = 0 = \Omega_0 C_0.$

That is, they are imposed in the form

$0 = \Omega_0 c_{0.1} = \Omega_0(H_{11}\gamma_{.1} + h_{.1}).$

It will be helpful to partition H_{11} and $h_{.1}$ to match the partitioning of Ω_0 (into Π_0 and I_{k_1}), as

$H'_{11} = (\mathbf{H}'_1 : \mathbf{H}'_2), \quad h'_{.1} = (\mathbf{h}'_1 : \mathbf{h}'_2)$;

then,

$0 = \Omega_0(H_{11}\gamma_{.1} + h_{.1})$

is equivalent to

$$(\Pi_0 : I_{k_1})\left\{\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix}\gamma_{.1} + \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix}\right\} = 0,$$

that is, to

$\Pi_0(\mathbf{H}_1\gamma_{.1} + \mathbf{h}_1) + (\mathbf{H}_2\gamma_{.1} + \mathbf{h}_2) = 0.$

The log-likelihood function will be maximised subject to this constraint, using the Lagrange multiplier $\tau_{2.1}$: for, it will be shown that the estimated Lagrange multiplier from this approach coincides with that obtained from the FIML approach.

Only the first-order conditions for vec Π_0 and $\gamma_{.1}$ will be obtained, since the objective is to establish that the "FIML" estimator $\tilde{\gamma}_{.1}$ and the Lagrange multiplier $\tilde{\tau}_{2.1}$ are obtained by this method. It is helpful to note here the identity

$$\mathbf{c}'\mathbf{D}\mathbf{f} = (\mathrm{vec}(\mathbf{c}\mathbf{f}'))'\mathrm{vec}\ \mathbf{D},$$

which follows from equation <1.6.1.2>.

The first-order condition for vec Π_0 is then

$$n^{-1}(\tilde{\Omega}_0^{-1} \otimes X')\mathrm{vec}(Y - X\tilde{\Pi}_0) + \mathrm{vec}(\tilde{\tau}_{2.1}(\mathbf{H}_1\tilde{\gamma}_{.1} + \mathbf{h}_1)') = 0,$$

or in non-vec form,

$$n^{-1}X'(Y - X\tilde{\Pi}_0)\tilde{\Omega}_0^{-1} + \tilde{\tau}_{2.1}(\mathbf{H}_1\tilde{\gamma}_{.1} + \mathbf{h}_1)' = 0, \qquad \text{<3.7.6.1>}$$

whilst the first-order condition for $\gamma_{.1}$ is

$$(\mathbf{H}_1'\tilde{\Pi}_0' + \mathbf{H}_2')\tilde{\tau}_{2.1} = H_{11}'\tilde{B}_0'\tilde{\tau}_{2.1} = 0. \qquad \text{<3.7.6.2>}$$

To obtain an expression for $\tilde{\tau}_{2.1}$, postmultiply the first-order condition for $\tilde{\Pi}_0$, equation <3.7.6.1>, by $\tilde{\Omega}_0\tilde{A}_0 e_1$ (where e_1 is the first m-dimensional coordinate vector), to yield

$$n^{-1}X'(Y - X\tilde{\Pi}_0)\tilde{A}_0 e_1 = -\tilde{\tau}_{2.1}(\mathbf{H}_1\tilde{\gamma}_{.1} + \mathbf{h}_1)'\tilde{\Omega}_0\tilde{A}_0 e_1. \qquad \text{<3.7.6.3>}$$

Because of the partitioning of B_0 and H_{11}, one can write

$$\mathbf{H}_1\tilde{\gamma}_{.1} + \mathbf{h}_1 = (\mathbf{I}_m : 0)(H_{11}\tilde{\gamma}_{.1} + h_{.1}) = (\mathbf{I}_m : 0)\tilde{c}_{0.1} = \mathbf{a} = \tilde{A}_0 e_1,$$

so that

$$(\mathbf{H}_1\tilde{\gamma}_{.1} + \mathbf{h}_1)'\tilde{\Omega}_0\tilde{A}_0 e_1 = \mathbf{a}'\tilde{\Omega}_0\mathbf{a} = \tilde{\sigma}_{0.11}.$$

Then, one can solve equation <3.7.6.3> to give

$$\begin{aligned}\tilde{\tau}_{2.1} &= -\tilde{\sigma}_{0.11}^{-1}n^{-1}X'(Y - X\tilde{\Pi}_0)\tilde{A}_0 e_1 \\ &= -\tilde{\sigma}_{0.11}^{-1}n^{-1}X'\tilde{V}_0\tilde{A}_0 e_1 \\ &= -\tilde{\sigma}_{0.11}^{-1}n^{-1}X'\tilde{u}_{0.1},\end{aligned}$$

which is precisely the expression given by equation <3.7.4.3>.

Substituting back into the first-order condition <3.7.6.2>, and recalling that

$$\tilde{u}_{0.1} = Z_1\tilde{c}_{0.1} = Z_1(H_{11}\tilde{\gamma}_{.1} + h_{.1}),$$

the LIML "normal equations" of equation <3.7.3.10> are obtained, which shows that

$$\tilde{\gamma}_{.1} = -[H_{11}'\tilde{\Pi}_0'X'Z_1H_{11}]^{-1}H_{11}'\tilde{\Pi}_0'X'Z_1h_{.1},$$

confirming the presumption that the "FIML" estimator of $\gamma_{.1}$ coincides with the Anderson and Rubin LIML estimator.

To obtain the limiting distribution of the Lagrange multiplier $\tilde{\tau}_{2.1}$ directly, note that the LIML residual vector for the first equation is defined by

$$\begin{aligned}\tilde{u}_{0.1} &= Z_1\tilde{c}_{0.1} = Z_1(H_{11}\tilde{\gamma}_{.1} + h_{.1})\\ &= -[I_n - Z_1H_{11}(H_{11}'\tilde{\Pi}_0'X'Z_1H_{11})^{-1}H_{11}'\tilde{\Pi}_0'X']Z_1h_{.1}.\end{aligned}$$

However, since

$$Z_1c_{0.1} = Z_1H_{11}\gamma_{.1} + Z_1h_{.1} = u_{0.1},$$

one can write

$$\tilde{u}_{0.1} = -[I_n - Z_1H_{11}(H_{11}'\tilde{\Pi}_0'X'Z_1H_{11})^{-1}H_{11}'\tilde{\Pi}_0'X']u_{0.1}.$$

Then,

$$X'\tilde{u}_{0.1} = P_{1n}'(\tilde{\alpha})X'u_{0.1}, \qquad \langle 3.7.6.4\rangle$$

where

$$P_{1n}(\tilde{\alpha}) = I_{k_1} - \tilde{\Pi}_0H_{11}(H_{11}'Z_1'X\tilde{\Pi}_0H_{11})^{-1}H_{11}'Z_1'X. \qquad \langle 3.7.6.5\rangle$$

As $n \to \infty$, $\tilde{\alpha} \xrightarrow{P} \alpha^0$, $\tilde{\Pi}_0 \xrightarrow{P} \Pi^0$, $n^{-1}Z_1'X \xrightarrow{P} \Pi^{0\prime}M_X$, and

$$P_{1n}(\tilde{\alpha}) \xrightarrow{P} P_1(\alpha^0) = I_{k_1} - \Pi^0H_{11}(H_{11}'\Pi^{0\prime}M_X\Pi^0H_{11})^{-1}H_{11}'\Pi^{0\prime}M_X,$$

so that

$$\begin{aligned}n^{1/2}\tilde{\tau}_{2.1} &\overset{a}{\approx} -\sigma_{.11}^{0-1}P_1'n^{-1/2}X'u_{0.1}\\ &\overset{a}{\approx} N(0,\ \sigma_{.11}^{0-1}P_1'M_XP_1),\end{aligned}$$

where $\sigma_{.11}^0$ is the (11) element of Σ^0.

3.7.7. The structural restrictions typically encountered in simultaneous equations models consist of within equation, exclusion restrictions and a unit normalisation rule: if this is to be true in the case of the single over-identified equation discussed in this section, the matrix H_{11} which imposes these restrictions in equation <3.7.2.3> will have to be block diagonal:

$$H_{11} = \begin{bmatrix} H_{11.1} & : & 0 \\ 0 & : & H_{11.2} \end{bmatrix},$$

as well as having a single non-zero element in each row and column. Similarly, the matrix H^*_{11} defined in equation <3.7.3.17> as

$$H^*_{11} = (h_{.1} : H_{11})$$

will have a block-diagonal structure, which may be expressed as

$$H^*_{11} = \begin{bmatrix} H^*_{11.1} & : & 0 \\ 0 & : & H_{11.2} \end{bmatrix}.$$

In order to relate the discussion in this section to the use of LIML estimation to provide the tests of identification described in the next Chapter, it will be helpful to rewrite the equation which generates the LIML estimator of $\gamma_{.1}$, equation <3.7.3.18>,

$$0 = H^{*\prime}_{11} Z'_1 (P_X - \nu I_n) Z_1 H^*_{11} \gamma^*_{.1},$$

to take account of the special structure of the restrictions imposed on the first equation.

It will also be convenient to make a transformation from the generalised characteristic root ν to

$$\iota^* = (1 - \nu)^{-1}\nu,$$

so that equation <3.7.3.18> becomes

$$0 = H_{11}^{*\prime}Z_1'[P_X - \iota^*(I_n - P_X)]Z_1H_{11}^*\gamma_{.1}^*; \qquad \langle 3.7.7.1\rangle$$

choosing the smallest root ι^* corresponds to choosing the smallest ν. Recalling that

$$Z_1 = (Y : X),$$

and temporarily putting

$$G = P_X - \iota^*(I_n - P_X),$$

equation <3.7.3.18> becomes

$$0 = \begin{bmatrix} H_{11.1}^{*\prime}Y'GYH_{11.1}^* & : & H_{11.1}^{*\prime}Y'XH_{11.2} \\ H_{11.2}'X'YH_{11.1}^* & : & H_{11.2}'X'XH_{11.2} \end{bmatrix}\gamma_{.1}^*.$$

The generalised characteristic root ι^* is a root of the determinental equation

$$0 = \det\begin{bmatrix} H_{11.1}^{*\prime}Y'GYH_{11.1}^* & : & H_{11.1}^{*\prime}Y'XH_{11.2} \\ H_{11.2}'X'YH_{11.1}^* & : & H_{11.2}'X'XH_{11.2} \end{bmatrix}$$

$$= \det(H_{11.2}'X'XH_{11.2})\det(H_{11.1}^{*\prime}Y'(G - P)Y\ H_{11.1}^*)$$

$$= \det[H_{11.1}^{*\prime}Y'((P_X - P) - \iota^*(I_n - P_X))YH_{11.1}^*], \qquad \langle 3.7.7.2\rangle$$

where

$$P = XH_{11.2}(H_{11.2}'X'XH_{11.2})^{-1}H_{11.2}'X'.$$

One can relate this even more directly to the existing literature by writing

$$Y_0 = YH_{11.1}^*,\quad X_1 = XH_{11.2},$$

producing the determinental equation

$$0 = \det Y_0'((P_X - P) - \iota^*(I_n - P_X))Y_0,$$

which is the form commonly found : see, for example, Fisk [1967, p43].

3.7.8. A two-step limited information estimator of the single over-identified equation

$Z_1 c_{0.1} = u_{0.1}$,

$c_{0.1} = H_{11}\gamma_{.1} + h_{.1}$

is easily obtained heuristically by analogy with equation <3.5.1.5> as

$$\hat{\hat{\gamma}}_{.1} = \gamma^{\#}_{.1} - (H_{11}'Q_0^{\#\prime}X'XQ_0^{\#}H_{11})^{-1}H_{11}'Q_0^{\#\prime}X'u_0^{\#}, \qquad \langle 3.7.8.1\rangle$$

$\gamma^{\#}_{.1}$ being some inefficient structural estimator, like the Khazzoom [1976] estimator, which solves by Moore-Penrose inverse, the equation

$\hat{Q}(H_{11}\gamma_{.1} + h_{.1}) = 0.$

In the expression <3.7.8.1> for $\hat{\hat{\gamma}}_{.1}$, $Q_0^{\#}$ involves the reduced form parameter estimator $\Pi_0^{\#}$ implied by the single over-identified equation model structure and $\gamma^{\#}_{.1}$. It is worth noting that if $\gamma^{\#}_{.1}$ is the two-stage least squares estimator $\hat{\gamma}_{.1}$, then

$\hat{\hat{\gamma}}_{.1} \equiv \hat{\gamma}_{.1}$,

simply because

$$\begin{aligned} H_{11}'\hat{Q}'X'\hat{u}_0 &= H_{11}'Z_1'P_X[I_n - Z_1H_{11}(H_{11}'Z_1'P_XZ_1H_{11})^{-1}H_{11}'Z_1'P_XZ_1]h \\ &= 0, \end{aligned}$$

where $\hat{u}_0$ is the two-stage least squares residual vector. This has also been observed by Hendry [1976]; the real reason is that two-stage least squares and LIML are asymptotically equivalent, but the two-step principle requires inefficient initial estimators.

3.8. Conclusions

3.8.1. The fundamental results on the nature of the maximum likelihood and related estimators of the parameters of a linear simultaneous equations model, and their statistical properties, have been established in this Chapter. Whilst the results obtained for the FIML and LIML estimators are basically the same as those obtained by Hendry [1976], the results for the minimum chi-squared estimator are believed to be new, and provide a useful link with more traditional estimators in the simultaneous equations model.

The full minimum chi-squared estimator is interesting, in that its criterion function, given in equation <3.6.1.1>, is extremely simple, although the minimum chi-squared estimates need to be obtained iteratively. The linearised minimum chi-squared estimator is based on one step of an iteration for the minimum chi-squared estimator: the linearity of this estimator creates a relationship with the well-known three-stage and indirect least squares estimators, although the linearised minimum chi-squared estimator based on efficient estimators of a just-identified model does have a characteristic form.

The discussion of two-step estimation shows that some care needs to be taken in constructing estimators by analogy with the form of the FIML estimator of the free structural parameters, as evidenced by the analysis of an assertion of

Hendry [1976] in subsections 3.5.2.-3.5.4..

The need for a discussion of LIML estimation in a manner compatible with the type of restrictions employed with the FIML estimator arises from the desire to discuss tests of over-identifying restrictions for single equations of a simultaneous equations model, as well as providing a basis for the discussion of tests of identification in the next Chapter. As seems usual with LIML estimation, the derivation is quite complex, even when obtained as a specialisation of the general FIML results to the case of an m-equation model containing only one over-identified equation. It is possible, however, to obtain the Lagrange multiplier associated with the structural parameter restrictions of this over-identified equation, by using the Anderson and Rubin [1949] type of approach, but the same difficulties arise in finding the estimator of $\gamma_{.1}$.

3.8.2. The results obtained in this Chapter are basic to the inferential results for the simultaneous equations model to be obtained in Chapters 4, 5, 6 and 9.

Chapter 4: Tests of the Conditions for Identification.

4.1. Introduction

4.1.1. The nature of the observationally equivalent structures of the null hypothesis model defined by equations <3.1.3.5> and <3.1.3.6>,

$(I_m \otimes Z_1)g_0 = u_0,$

$g_0 = H\gamma + h,$

where Z_1 is defined by equation <3.1.2.3>,

$Z_1 = (Y : X),$

can be deduced from the discussion of the alternative hypothesis model (defined by equations <3.1.2.6> and <3.1.2.7>) in subsection 3.2.2.: it then follows that a structure associated with a specific value of the parameter vector γ is identified if and only if γ is the unique solution to

$$(I_m \otimes \Omega_0)(H\gamma + h) = 0, \quad \langle 4.1.1.1\rangle$$

where

$\Omega_0 = (\Pi_0 : I_{k_1}),$

and Π_0 is the reduced form parameter value implied by the specific value of γ. This condition can be rephrased, as in subsection 3.2.3., to say that γ is identified if and only if there exists a solution to equation <4.1.1.1>, and

$$\text{rank } (I_m \otimes \Omega_0)H = q_0: \quad \langle 4.1.1.2\rangle$$

recall from subsection 3.1.3. that the matrix H has q_0 linearly independent columns. These conditions are referred to as the "consistency condition" and the "rank condition" respectively.

In this Chapter, some tests of the rank condition are devised, followed by the construction of a similar test of the consistency condition; these tests require the derivation of the limit distribution of the characteristic roots of certain random symmetric matrices, based on fundamental arguments given by Anderson [1963].

4.1.2. Before proceeding to an outline of the contents of this Chapter, it is useful to consider the reasons why tests of the rank and consistency conditions for identification might be helpful and informative to an investigator.

Given the choice of endogenous and exogenous variables, the investigator's substantive economic theory is encapsulated in the structural restrictions
$g_0 = H\gamma + h$:
if distinct behavioural reactions, described by the value of the parameter vector γ, do not lead to distinct observable features of the distribution of the endogenous variables y_t, so that different values of γ lead to the same values of Π_0 and Ω_0, then the model-building process will fail in its objective of describing and explaining behavioural reactions. To the extent that particular elements of the vector γ are identified (i.e. have the same values) within the set of observationally equivalent structures satisfying
$g_0 = H\gamma + h$,

this criticism is mitigated.

Conceptually, inference on "true" parameters whose values are not unique is difficult to imagine, and there is a corresponding "sample" difficulty, which is the equivalent of multicollinearity. The limit covariance matrix of the FIML estimator $\tilde{\gamma}$ was given in equation <3.4.3.11> as

$$[H'(\Sigma^{0-1} \otimes \Omega^{0\prime}M_x\Omega^0)H]^{-1}:$$

this inverse will not exist if

$$(I_m \otimes \Omega^0)H$$

does not have full column rank, which will occur if there is a failure of identification at the true parameter value (under the null hypothesis). It will also be seen in Chapter 6 that the degrees of freedom of test statistics for tests of overidentifying restrictions are affected, leading to incorrect inferences if the wrong degrees of freedom are used. The traditional consequence of multicollinearity in linear regression, "large" standard errors of the estimated parameters, will also tend to occur.

In addition, if

$$(I_m \otimes \tilde{\Omega}_0)H$$

should have less than full column rank at any stage in the iterative process of estimation, it is unlikely that the maximum of the log-likelihood function will be found, unless special precautions (like the use of ridge regression-type corrections to the second derivative or information matrices) are taken.

It would therefore seem desirable to have a statistical test for the possibility that

$(I_m \otimes \Omega_0)H$

has a rank failure at the true parameter point, preferably before the complex apparatus of FIML estimation has been set in motion. However, one could argue that with the accuracy of modern numerical methods, the sample matrix

$(I_m \otimes \tilde{\Omega}_0)H$

may appear to have full column rank, even though

$(I_m \otimes \Omega^0)H$

has less than full column rank: a test of this possibility is again desirable.

<u>4.1.3.</u> In the next section, the existing literature on tests of identification is surveyed: this literature has only considered the problem in the context of the LIML estimation of a single structural equation from a simultaneous equations model, in fact, a special case of the situation analysed in section 3.7. . The literature is described in terms of the notation used in that section, to permit a satisfactory comparison with the results obtained in this Chapter.

The next section discusses the choice of criteria for a test of the rank condition, followed by a derivation of certain limit distributions for the characteristic roots of a symmetric matrix which in turn generates the limit distribution of the test statistics. These results are then specialised to deal with the "usual special case" of the simultaneous equations model in which there are only within equation exclusion restrictions and a unit normalisation rule. Joint and sequential tests of rank are then considered; the final section considers a test of the consistency condition, based on the same arguments as the proposed test of the rank condition.

So far as the author is aware, the proposed test statistics and test procedures have not yet appeared in the literature.

4.2. Tests of Identification - A Survey.

4.2.1. The existing literature on tests of identification uses the LIML estimator exclusively, and for the special case where the restrictions on the over-identified equation considered are exclusion restrictions, and usually, but not always, a unit normalisation rule. This literature will be discussed using the notation associated with the more general restrictions employed in section 3.7., to maintain comparability, although this notation is not to be found in the works referred to. In addition, attention will be focussed on the roots of determinental equations like

$$\det H_{11}^{*\prime} Z_1^{\prime} [P_X - \iota^*(I_n - P_X)] Z_1 H_{11}^* = 0, \qquad \text{<4.2.1.1>}$$

arising from the version of equation <3.7.3.18>,

$$0 = H_{11}^{*\prime} Z_1^{\prime} [P_X - \iota^*(I_n - P_X)] Z_1 H_{11}^* \gamma_{.1}^*, \qquad \text{<4.2.1.2>}$$

given as equation <3.7.7.1> of subsection 3.7.7. .

It will be convenient to rewrite equation <4.2.1.2> above slightly: note that

$$n^{-1} Z_1^{\prime}(I_n - P_X) Z_1 = \begin{bmatrix} n^{-1} Y^{\prime}(I_n - P_X) Y & : & 0 \\ 0 & : & 0 \end{bmatrix}$$

and that

$$n^{-1} Y^{\prime}(I_n - P_X) Y$$

is the unrestricted least squares estimator $\hat{\Omega}$ of the reduced form covariance matrix Ω_0. Then, one can write

$$n^{-1} H_{11}^{*\prime} Z_1^{\prime} P_X Z_1 H_{11}^* \gamma_{.1}^* = \iota^* H_{11}^{*\prime} \begin{bmatrix} \hat{\Omega} & : & 0 \\ 0 & : & 0 \end{bmatrix} H_{11}^* \gamma_{.1}^*, \qquad \text{<4.2.1.3>}$$

and the LIML estimator $\tilde{\gamma}_{.1}^*$ arises as the characteristic vector associated with the smallest generalised characteristic root $\tilde{l}^*$.

Certain statistics based on the two smallest generalised characteristic roots of equation <4.2.1.3> have traditionally been used to provide "tests of identification", but the precise nature of the hypotheses being tested by such statistics has not always been made clear, and this has been the source of some confusion in the literature. These tests stem originally from the work of Anderson and Rubin [1949, 1950] and Koopmans and Hood [1953], using LIML estimation, and have been re-examined by Farebrother and Savin [1974] in the context of a general k-class estimator. The original work on the "single root" and "double root" tests mentioned below is fully discussed in Fisk [1967], particularly with reference to similar situations in the multivariate analysis literature.

4.2.2. Rather than simply describe the various test statistics, the null and alternative hypotheses for which they are designed are considered, together with the relationship of these hypotheses to the conditions for the parameter vector $\gamma_{.1}$ to be identified.

For this purpose, it is helpful to consider a population analogue to equation <4.2.1.3>, namely,

$$n^{-1}H_{11}^{*\prime}\Pi_0^{\prime}X^{\prime}X\Pi_0H_{11}^{*}\gamma_{.1}^{*} = l^{*}H_{11}^{*\prime}\begin{bmatrix}\Omega_0 & : & 0\\ 0 & : & 0\end{bmatrix}H_{11}^{*}\gamma_{.1}^{*}, \qquad \langle 4.2.2.1\rangle$$

(compare Farebrother and Savin [1974, equation 19]), and also the set of equations obtained by deleting the first, linearly dependent, row of $H_{11}^{*\prime}$:

$$n^{-1}H_{11}^{\prime}\Pi_0^{\prime}X^{\prime}X\Pi_0(H_{11}\gamma_{.1} + h_{.1}) = lH_{11}^{\prime}\begin{bmatrix}\Omega_0 & : & 0\\ 0 & : & 0\end{bmatrix}(H_{11}\gamma_{.1} + h_{.1}), \qquad \langle 4.2.2.2\rangle$$

with generalised characteristic root l.

Denote the ordered generalised characteristic roots associated with <4.2.2.1> by

$l_1^{*} \geq \ldots \geq l_{q_{01}+1}^{*} \geq 0$,

and of equation <4.2.2.2> by

$l_1 \geq \ldots \geq l_{q_{01}} \geq 0$:

then, it is known (see Fisk [1967, p49]) that a solution $\gamma_{.1}^{**}$ to <4.2.2.1> exists if and only if

$l_{q_{01}+1}^{*} = 0$,

and that the solution for $\gamma_{.1}$ from <4.2.2.2> is unique (i.e. $\gamma_{.1}$ is identified) if and only if in addition

$l_{q_{01}} > 0$

(see Farebrother and Savin [1974, p382]). Furthermore, the generalised characteristic roots l_i^{*}, l_i satisfy a Sturmian Separation Theorem (see Farebrother [1974]), namely,

$$l_{q_{01}+1}^{*} \leq l_{q_{01}} \leq l_{q_{01}}^{*} \leq l_{q_0-1} \leq \ldots \leq l_1 \leq l_1^{*}. \qquad \langle 4.2.2.3\rangle$$

The conventional approach to the single root and double root tests is described by Fisk [1967], for example: a test of the hypothesis

H_0: $\quad l^*_{q_{01}+1} = 0$ <4.2.2.4>

H_1: $\quad l^*_{q_{01}+1} > 0$ <4.2.2.5>

uses the smallest root $\tilde{l}^*$ of the determinental equation <4.2.1.1>. If the null hypothesis is thereby rejected, then one may say, as does Fisk [1967], that there is evidence of a lack of identifiability. If the null hypothesis is accepted, then one proceeds to a "double root" test of the hypotheses

H^*_0: $\quad l^*_{q_{01}+1} = l^*_{q_{01}} = 0$

H^*_1: $\quad l^*_{q_{01}+1} = 0, \; l^*_{q_{01}} > 0,$

the test statistic being based on the two smallest roots of the equation <4.2.1.1>. If this null hypothesis is rejected, then one concludes that there is evidence that the equation (i.e. $\gamma_{.1}$) is identified.

However, from the interlacing of the characteristic roots l^*_i, l_i described in equation <4.2.2.3>, it can be seen that the assertion

$l^*_{q_{01}} > 0$

is not equivalent to the assertion that

$l_{q_{01}} > 0,$

which is the condition required for the identification of $\gamma_{.1}$, given that $l^*_{q_{01}+1} = 0$. Consequently, rejection of the null hypothesis H^*_0 of the double root test is not equivalent to acceptance of the hypothesis that $\gamma_{.1}$ is identified.

This argument leads naturally to the conclusion that it would be better to test $l_{q_{01}} > 0$ directly: this conclusion was reached by Farebrother and Savin [1974], who also provided a suitable test statistic (Farebrother and Savin [1974, equation 21]). The null hypothesis is

$$H_0^{*}: \quad \ell_{q_{01}} = 0, \qquad \langle 4.2.2.6\rangle$$

with alternative

$$H_1^{*}: \quad \ell_{q_{01}} > 0; \qquad \langle 4.2.2.7\rangle$$

the test statistic is based on the smallest root $\tilde{\ell}$ of the determinental equation

$$\det H_{11}'Z_1'(P_X - \ell(I_n - P_X))Z_1H_{11} = 0$$

analogous to equation <4.2.1.1>. The test itself is based on the limit distribution

$$n \log \tilde{\ell} \overset{a}{\approx} \chi^2_{k_1-q_{01}+1}:$$

in Farebrother and Savin's [1974] notation, the degrees of freedom are $k_2 - g + 1$.

The null hypothesis of this test is clearly one of "failure of identification"; given that H_0^{*} is false, one should then proceed to the traditional single root test of the hypotheses of equations <4.2.2.4> and <4.2.2.5>, using the smallest root $\tilde{\ell}^*$ of the determinental equation <4.2.1.1>. If the null hypothesis of equation <4.2.2.4> is then accepted, one can interpret the results as saying "there exists a unique $\gamma_{.1}$ satisfying equation <4.2.2.2>."

If the structural equation being considered is just identified, this last assertion is an automatic consequence of the rank condition, from equation <4.2.2.2>: thus, the hypothesis

$$H_0: \quad \ell^*_{q_{01}+1} = 0$$

is only falsifiable in the case of an over-identified equation. In such an equation, the truth of this null hypothesis is equivalent to the truth of the restrictions of equation <3.7.2.3>,

$c_{0.1} = H_{11}\gamma_{.1} + h_{.1}$,

and hence to a test of the "over-identifying restrictions" of <3.7.3.2> against any just-identified alternative, for it is well known (see Fisk [1967,pp51,53]) that the "single root" test statistic

$n \log \tilde{l}^* \overset{a}{\sim} \chi^2_{k_1-q_{01}}$

is asymptotically equivalent, under the null hypothesis just stated, to a test of the overidentifying restrictions based on any asymptotically efficient limited information estimator.

4.3. Sample Criteria for Identification

4.3.1. If one accepts the arguments given in subsection 4.1.2. concerning the need for a test of the rank condition of equation <4.1.1.2>,

rank $(I_m \otimes \Omega_0)H = q_0$,

(i.e. full column rank), then it will be necessary to propose and analyse some test statistics, which, so far as the author is aware, have not appeared in the literature. Ideally, these tests should not be dependent on a specific choice of estimator, or on the examination of the equations of the model one at a time, as is the case with the existing tests described in the preceding section.

The method proposed in this thesis for generating such tests is to use the smallest roots of estimates of matrices of the form

$$H'(I_m \otimes \Omega_0')N(I_m \otimes \Omega_0)H, \qquad \text{<4.3.1.1>}$$

where choice may be made both of the estimator of Π_0 in

$\Omega_0 = (\Pi_0 : I_{k_1})$

and of the positive definite matrix N. Such a matrix as <4.3.1.1> will be described as a "criterion matrix": one possible choice is a finite sample version of the limiting covariance matrix of the maximum likelihood estimator $\tilde{\gamma}$ of the null hypothesis model defined by equations <3.1.3.5> and <3.1.3.6>,

$$H'(\Sigma_0^{-1} \otimes \Omega_0'X'X\Omega_0)H. \qquad \text{<4.3.1.2>}$$

Nonsingularity of this matrix corresponds to Rothenberg's

[1971] criterion that the parameter vector γ is identified if and only if the "finite sample" information matrix is nonsingular. This condition is also appropriate even in the case where the matrix X does not have full column rank, so that some of the restrictions in equation <3.1.3.6>,

$g_0 = H\gamma + h$

serve to "correct" the rank deficiency in $(I_m \otimes X\Omega_0)$.

A test statistic can therefore be obtained from the smallest characteristic roots of an estimate,

$$H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Omega}_0' X' X \tilde{\Omega}_0)H \qquad \text{<4.3.1.3>}$$

of the matrix <4.3.1.2>: here, the FIML estimators are being used. It will be argued later that the null hypothesis of a test of identification should be that the model or parameter vector γ is identified, and thus, one can expect that the FIML estimators of γ, Π_0, Ω_0 and Σ_0 will exist, so that the test being proposed may be described as a "post-estimation" test of identification.

However, it is possible that a population rank failure will be deemed to have occurred by the hypothesis test, even though the smallest roots of <4.3.1.3> are not so small as to prevent numerically accurate structural parameter estimation. This does not involve any contradiction, since it is merely a tribute to the accuracy of modern numerical methods in the presence of ill-conditioning.

<u>4.3.2.</u> It may also be argued that if there is a failure of

the rank condition, it may be possible, and indeed desirable, to detect this prior to "expensive" structural estimation. With this in mind, it is proposed to use as a criterion matrix

$$H'(I_m \otimes \hat{Q}'X'X\hat{Q})H, \qquad \text{<4.3.2.1>}$$

that is, using the unrestricted reduced form estimator $\hat{\pi}$. One possible advantage of such a "preliminary" test of identification is that a more refined estimator of π_0, like the FIML estimator $\tilde{\pi}_0$, may be numerically more susceptible to near lack of identification than a non-structural estimator like the Ordinary Least Squares estimator $\hat{\pi}$, making it more difficult to actually obtain the FIML estimates for the post-estimation test of identification.

Another significant point is that the limit distribution of the characteristic roots of the post-estimation criterion <4.3.1.3> depends on the assumption that the restrictions <3.1.3.6>

$$g_0 = H\gamma + h$$

are true, whereas limiting distribution of the characteristic roots of the preliminary test criterion <4.3.2.1> essentially depend only on the limiting distribution of the unrestricted reduced form estimator $\hat{\pi}$, and are thus independent of the truth of the restrictions <3.1.3.6>.

This particular approach of using the characteristic roots of sample criterion matrices is quite useful in the sense that any estimator which is asymptotically equivalent to the FIML estimator can be used in <4.3.1.3>. For estimators not asymptotically equivalent to FIML, the limiting distributions of the distinct elements of the sample criterion matrices need to be modified appropriately.

The intuitive argument for conducting such tests can be summarised by saying that models which fail such tests of identification are not specified "tightly" enough with respect to the data set being used, and are thus suffering from a misspecification which it may be possible to correct.

4.4. Limit Distributions of Sample Rank Criterion Matrices

4.4.1. In order to obtain limit distributions for the characteristic roots of the sample rank criterion matrices <4.3.1.3> and <4.3.2.1>, it will be necessary to obtain limit normal distributions for their distinct elements: for this purpose, denote the population criterion matrix <4.3.1.2> by

$$R_n(\pi_0) = n^{-1}H'(\Sigma_0^{-1} \otimes \Pi_0'X'X\Pi_0)H \quad \text{<4.4.1.1>}$$

and the population analogue of <4.3.2.1> by

$$R_n^*(\pi_0) = n^{-1}H'(I_m \otimes \Pi_0'X'X\Pi_0)H. \quad \text{<4.4.1.2>}$$

The sample matrices <4.3.1.3> and <4.3.2.1> are then

$$R_n(\tilde{\pi}_0) = n^{-1}H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'X\tilde{\Pi}_0)H \quad \text{<4.4.1.3>}$$

$$R_n^*(\hat{\pi}) = n^{-1}H'(I_m \otimes \hat{\Pi}'X'X\hat{\Pi})H, \quad \text{<4.4.1.4>}$$

and it will also be convenient to use the matrix

$$R_n^*(\tilde{\pi}_0) = n^{-1}H'(\Sigma_0^{-1} \otimes \tilde{\Pi}_0'X'X\tilde{\Pi}_0)H.$$

Although $R_n(\tilde{\pi}_0)$ depends on $\tilde{\Sigma}_0$, this dependence is not important for the arguments that follow. For, a Taylor series expansion argument can be used to obtain the limit distribution of the distinct elements,

$v[R_n(\tilde{\pi}_0)]$

of $R_n(\tilde{\pi}_0)$ and

$v[R_n^*(\hat{\pi})]$

of $R_n^*(\hat{\pi})$ in terms of the known limiting distributions of

$n^{1/2}(\tilde{\pi}_0 - \pi^0) = n^{1/2}\text{vec }(\tilde{\Pi}_0 - \Pi^0)$,

$n^{1/2}(\hat{\pi} - \pi^0) = n^{1/2}\text{vec }(\hat{\Pi} - \Pi^0)$.

To obtain the necessary derivatives, a differential argument will be utilised, in which the results of subsections 1.6.2. and 1.6.3. will be freely used. Thus,

$$dR_n(\pi_0) = n^{-1}H'[\Sigma_0^{-1} \otimes d(\Omega_0'X'X\Omega_0)]H$$
$$= n^{-1}H'[\Sigma_0^{-1} \otimes (d\Omega_0)'X'X\Omega_0]H + n^{-1}H'[\Sigma_0^{-1} \otimes \Omega_0'X'Xd\Omega_0]H,$$

so that

$$\mathrm{dvec}\ R_n(\pi_0) = \mathrm{vec}\ dR_n(\pi_0)$$
$$= n^{-1}[H'(\Sigma_0^{-1} \otimes \Omega_0'X'X\Omega_0) \otimes H']\mathrm{vec}(\mathbf{I}_m \otimes (d\Omega_0)')$$
$$+ n^{-1}[H' \otimes H'(\Sigma_0^{-1} \otimes \Omega_0'X'X)]\mathrm{vec}(\mathbf{I}_m \otimes d\Omega_0).$$

By the results <1.6.3.1> and <1.6.3.3>,

$$\mathrm{vec}(\mathbf{I}_m \otimes (d\Omega_0)') = \mathbf{K}_{m(m+k_1),mk_1}\mathrm{vec}(\mathbf{I}_m \otimes d\Omega_0)$$

and

$$n^{-1}[H'(\Sigma_0^{-1} \otimes \Omega_0'X'X) \otimes H']\mathbf{K}_{m(m+k_1),mk_1}$$
$$= \mathbf{K}_{q_0}[H' \otimes H'(\Sigma_0^{-1} \otimes \Omega_0'X'X)],$$

so that

$$\mathrm{vec}\ dR_n(\pi_0) = n^{-1}(\mathbf{I}_{q_0^2} + \mathbf{K}_{q_0})[H' \otimes H'(\Sigma_0^{-1} \otimes \Omega_0'X'X)]\mathrm{vec}(\mathbf{I}_m \otimes d\Omega_0)$$
$$= 2n^{-1}\mathbf{S}_{q_0}[H' \otimes H'(\Sigma_0^{-1} \otimes \Omega_0'X'X)]\mathrm{vec}(\mathbf{I}_m \otimes d\Omega_0),$$

using the definition <1.6.3.5> of $\mathbf{S_n}$.

Now,

$$\Omega_0 = (\Pi_0 : \mathbf{I}_{k_1}),$$

so that

$$d\Omega_0 = (d\Pi_0 : 0) = d\Pi_0(\mathbf{I}_m : 0)$$

and

$$\mathrm{dvec}\ \Omega_0 = \left[\begin{bmatrix}\mathbf{I}_m \\ 0\end{bmatrix} \otimes \mathbf{I}_{k_1}\right]\mathrm{vec}\ d\Pi_0$$
$$= \begin{bmatrix}\mathbf{I}_{mk_1} \\ 0_{k_1^2,mk_1}\end{bmatrix}\mathrm{vec}\ d\Pi_0$$
$$= F\mathrm{vec}\ d\Pi_0,$$

say.

Then, using equation <1.6.4.3>,

$$\mathrm{vec}(I_m \otimes dQ_0) = \begin{bmatrix} I_{m+k_1} \otimes e_1 \otimes I_{k_1} \\ \cdot \\ \cdot \\ I_{m+k_1} \otimes e_m \otimes I_{k_1} \end{bmatrix} \mathrm{vec}\, dQ_0$$

$$= \begin{bmatrix} I_{m+k_1} \otimes e_1 \otimes I_{k_1} \\ \cdot \\ \cdot \\ I_{m+k_1} \otimes e_m \otimes I_{k_1} \end{bmatrix} \left[\begin{bmatrix} I_m \\ 0 \end{bmatrix} \otimes 1 \otimes I_{k_1} \right] \mathrm{vec}\, d\Pi_0$$

$$= \begin{bmatrix} \begin{bmatrix} I_m \\ 0 \end{bmatrix} \otimes e_1 \otimes I_{k_1} \\ \cdot \\ \cdot \\ \begin{bmatrix} I_m \\ 0 \end{bmatrix} \otimes e_m \otimes I_{k_1} \end{bmatrix} \mathrm{vec}\, d\Pi_0, \qquad \langle 4.4.1.5 \rangle$$

where e_i, $i = 1,\ldots,m$ are m-dimensional coordinate vectors.

To complete the derivative expression for vec $dR_n(\pi_0)$, let H be decomposed by row blocks, each having $m + k_1$ rows:

$$H' = (H_1',\ldots,H_m'), \qquad \langle 4.4.1.6 \rangle$$

so that one can write

$$[H' \otimes H'(\Sigma_0^{-1} \otimes Q_0'X'X)] = [H_1' \otimes H'(\Sigma_0^{-1} \otimes Q_0'X'X),\ldots,$$
$$H_m' \otimes H'(\Sigma_0^{-1} \otimes Q_0'X'X)]. \qquad \langle 4.4.1.7 \rangle$$

Note that in turn, each block H_i' may be further partitioned into m and k_1 rows:

$$H_i' = (H_{i.1}' : H_{i.2}'), \qquad \langle 4.4.1.8 \rangle$$

so that

$$H_i'\begin{bmatrix} I_m \\ 0 \end{bmatrix} = H_{i.1}'.$$

One can then premultiply the large matrix in equation <4.4.1.5> by the matrix in equation <4.4.1.7> to obtain

$n^{-1}\sum_{i=1}^{m}[H_{i.1}' \otimes H'(\Sigma_0^{-1}e_i \otimes \Omega_0'X'X)]$:

that is,

$$\text{dvec } R_n(\pi_0) = 2n^{-1}\mathbf{S}_{\mathbf{q}_0}\sum_{i=1}^{m}[H_{i.1}' \otimes H'(\Sigma_0^{-1}e_i \otimes \Omega_0'X'X)]\text{vec } d\Pi_0.$$

Finally, the relationship <1.6.3.9> between $v[R_n(\pi_0)]$ and vec $R_n(\pi_0)$ yields

$$dv[R_n(\pi_0)] = 2n^{-1}\mathbf{L}_{\mathbf{q}_0}\mathbf{S}_{\mathbf{q}_0}\sum_{i=1}^{m}[H_{i.1}' \otimes H'(\Sigma_0^{-1}e_i \otimes \Omega_0'X'X)]\text{vec } d\Pi_0.$$

It will then follow from the first-order Taylor series expansion

$$v[R_n^{\#}(\tilde{\pi}_0)] = v[R_n(\pi^0)] + D_{\pi_0}[v[R_n(\pi_0^*)]](\tilde{\pi}_0 - \pi^0),$$

$\pi_0^* \in (\tilde{\pi}_0, \pi^0)$, that

$$n^{1/2}v[R_n^{\#}(\tilde{\pi}_0) - R_n(\pi^0)] \overset{a}{\approx} 2\mathbf{L}_{\mathbf{q}_0}\mathbf{S}_{\mathbf{q}_0}\sum_{i=1}^{m}[H_{i.1}' \otimes H'(\Sigma^{0-1}e_i \otimes \Omega^{0\prime}X'X)] \times n^{1/2}(\tilde{\pi}_0 - \pi^0); \qquad \text{<4.4.1.9>}$$

since

$$\tilde{\Sigma}_0 \xrightarrow{P} \Sigma^0,$$

it follows that

$n^{1/2}v[R_n(\tilde{\pi}_0) - R_n(\pi^0)]$ and $n^{1/2}v[R_n^{\#}(\tilde{\pi}_0) - R_n(\pi^0)]$

have the same limit distribution. Thus, from the limit distribution statement <3.4.3.12> and <3.4.3.13>,

$$n^{1/2}(\tilde{\pi}_0 - \pi^0) \overset{a}{\approx} N(0, \Psi(\tilde{\pi}_0;\theta^0)),$$

with

$$\Psi(\tilde{\pi}_0;\theta^0) = (A^{0-1\prime} \otimes \Omega^0)H[H'(\Sigma^{0-1} \otimes \Omega^{0\prime}M_X\Omega^0)H]^{-1}H'(A^{0-1} \otimes \Omega^{0\prime});$$

the limit distribution of

$n^{1/2}v[R_n(\tilde{\pi}_0) - R_n(\pi^0)]$

can then be obtained from <4.4.1.9> as

$$n^{1/2}v[R_n(\tilde{\pi}_0) - R_n(\pi^0)] \overset{a}{\sim} N(0,\Psi[v[R_n(\tilde{\pi}_0)];\theta^0]) \qquad \text{<4.4.1.10>}$$

with

$$\Psi[v[R_n(\tilde{\pi}_0)];\theta^0] = 4\mathbf{L}_{\mathbf{q}_0}\mathbf{S}_{\mathbf{q}_0}\sum_{i=1}^{m}\sum_{j=1}^{m}[H'_{i.1}A^{0-1\prime} \otimes H'(\Sigma^{0-1}e_i \otimes B^{0\prime}M_x \times B^0)]H_i[H'(\Sigma^{0-1} \otimes B^{0\prime}M_xB^0)H]^{-1}H'[A^{0-1}H_{j.1} \otimes (e'_j\Sigma^{0-1} \otimes B^{0\prime}M_xB^0)H]\mathbf{S}_{\mathbf{q}_0}\mathbf{L}'_{\mathbf{q}_0}. \qquad \text{<4.4.1.11>}$$

<u>4.4.2.</u> The corresponding limit normal distribution for

$n^{1/2}v[R^*_n(\hat{\pi}) - R^*_n(\pi^0)]$

can be obtained from these results, simply by replacing Σ^0 by $\mathbf{I}_m$, and by using the limiting distribution of

$n^{1/2}\text{vec}\ (\hat{\Pi} - \Pi^0) = n^{1/2}(\hat{\pi} - \pi^0)$

given in equations <3.5.3.6> and <3.5.3.7>:

$$n^{1/2}(\hat{\pi} - \pi^0) \overset{a}{\sim} N(0,\Psi(\hat{\pi};\theta^0))$$

where

$$\Psi(\hat{\pi};\theta^0) = \Omega^0 \otimes M_x^{-1}.$$

This will yield

$$n^{1/2}v[R^*_n(\hat{\pi}) - R^*_n(\pi^0)] \overset{a}{\sim} N[0,\Psi(R^*_n(\hat{\pi});\theta^0)] \qquad \text{<4.4.2.1>}$$

with

$$\begin{aligned}\Psi(R^*_n(\hat{\pi});\theta^0) &= 4\mathbf{L}_{\mathbf{q}_0}\mathbf{S}_{\mathbf{q}_0}\sum_{i=1}^{m}\sum_{j=1}^{m}[H'_{i.1} \otimes H'(e_i \otimes B^{0\prime}M_x)] \\ &\qquad \times(\Omega^0 \otimes 1 \otimes M_x^{-1})[H_{j.1} \otimes (e'_j \otimes M_xB^0)H]\mathbf{S}_{\mathbf{q}_0}\mathbf{L}'_{\mathbf{q}_0} \\ &= 4\mathbf{L}_{\mathbf{q}_0}\mathbf{S}_{\mathbf{q}_0}\sum_{i=1}^{m}\sum_{j=1}^{m}[H'_{i.1}\Omega^0H_{j.1} \otimes H'(e_ie'_j \otimes B^{0\prime}M_xB^0)H] \\ &\qquad \times \mathbf{S}_{\mathbf{q}_0}\mathbf{L}'_{\mathbf{q}_0}. \end{aligned} \qquad \text{<4.4.2.2>}$$

4.4.3. As noted in section 4.2., the tests of identification proposed in the existing literature have been for the case of a single equation embodying only exclusion restrictions and a unit normalisation rule, unlike the "system" or joint tests proposed here. It is therefore useful to see how the results obtained above specialise in this case, if only for the purposes of comparison.

It will be necessary to examine the nature of the null hypothesis simultaneous equations model of <3.1.3.5> and <3.1.3.6> in this case:

$(I_m \otimes Z_1)g_0 = u_0$,

$g_0 = H\gamma + h$.

Because there are only within equation restrictions,

$H = \|H_{ij}\| = \oplus_{i=1}^{m} H_{ii}$:

that is, $H_{ij} = 0$ for $i \neq j$; each diagonal block is $(m + k_1) \times q_{0i}$, with

$q_0 = \sum_{i=1}^{m} q_{0i}$.

With this particular structure, there does not appear to be any simplification in the structure of $R_n(\tilde{\pi})$ or its limit distribution; however, the criterion $R_n^*(\hat{\pi})$ now becomes block diagonal:

$$R_n^*(\hat{\pi}) = n^{-1}H'(I_m \otimes \hat{\Omega}'X'X\hat{\Omega})H$$
$$= n^{-1} \oplus_{i=1}^{m} H_{ii}'\hat{\Omega}'X'X\hat{\Omega}H_{ii} = \oplus_{i=1}^{m} R_{in}^*(\hat{\pi}),$$

with the corresponding population analogue

$$R_{in}^*(\pi^0) = n^{-1}H_{ii}'\Omega^{0\prime}X'X\Omega^0 H_{ii}, \qquad \text{<4.4.3.1>}$$

$$R_n^*(\pi^0) = \oplus_{i=1}^{m} R_{in}^*(\pi^0). \qquad \text{<4.4.3.2>}$$

It is in fact easier to examine the impact of this structure on the limit distribution directly, making use of equation <1.6.4.2>. It follows from arguments similar to those used above to obtain the limit distribution of

$n^{½}v[R_n(\tilde{\pi}) - R_n(\pi^0)]$ and $n^{½}v[R_n^*(\hat{\pi}) - R_n^*(\pi^0)]$

that

$$n^{½}\text{vec}[R_{in}^*(\hat{\pi}) - R_{in}^*(\pi^0)] \overset{a}{\approx} 2S_{q_{0i}}(H'_{ii.1} \otimes H'_{ii}\Omega^{0\prime}M_X)n^{½}\text{vec}(\hat{\Pi}-\Pi^0),$$

$$i = 1,\ldots,m, \qquad \langle 4.4.3.3\rangle$$

where

$H'_{ii} = [H'_{ii.1} : H'_{ii.2}]$

is a partition of the $q_{0i} \times (m + k_1)$ matrix H'_{ii} into m and k_1 columns. Let E_i be the matrix defined in equation <1.6.4.1>:

$$E_i = \begin{bmatrix} 0_{q_{01},q_{0i}} \\ \cdot \\ \cdot \\ I_{q_{0i}} \\ \cdot \\ \cdot \\ 0_{q_{0m},q_{0i}} \end{bmatrix}.$$

so that, by <1.6.4.2>.

$$n^{½}\text{vec}[R_n^*(\hat{\pi})-R_n^*(\pi^0)] \overset{a}{\approx} \underset{i=1}{\overset{m}{\oplus}}(I_{q_{0i}} \otimes E_i)n^{½}\begin{bmatrix} \text{vec}[R_{1n}^*(\hat{\pi})-R_{1n}^*(\pi^0)] \\ \cdot \\ \cdot \\ \text{vec}[R_{mn}^*(\hat{\pi})-R_{mn}^*(\pi^0)] \end{bmatrix}$$

$$\langle 4.4.3.3\rangle$$

$$\overset{a}{\sim} 2 \bigoplus_{i=1}^{m}(I_{q_{0i}} \otimes E_i)\begin{bmatrix} S_{q_{01}}(H'_{11.1} \otimes H'_{11}\Omega^{0\prime}M_x) \\ \cdot \\ \cdot \\ S_{q_{0m}}(H'_{mm.1} \otimes H'_{mm}\Omega^{0\prime}M_x)\end{bmatrix} n^{1/2}\mathrm{vec}(\hat{\Pi} - \Pi^0).$$

From this, and the result that

$$n^{1/2}v[R_n^*(\hat{\pi}) - R_n^*(\pi^0)] = n^{1/2}\mathbf{L}_{q_0}\mathrm{vec}(R_n^*(\hat{\pi}) - R_n^*(\pi^0)),$$

one can establish that

$$n^{1/2}v[R_n^*(\hat{\pi}) - R_n^*(\pi^0)] \overset{a}{\sim} N(0,\Psi[v(R_n^*(\hat{\pi}));\theta^0]),$$

where

$$\Psi[v(R_n^*(\hat{\pi}));\theta^0] = \mathbf{L}_{q_0}AB A'\mathbf{L}'_{q_0},$$

and

$$A = \bigoplus_{i=1}^{m}(I_{q_{0i}} \otimes E_i),$$

$$B = 4\|S_{q_{0i}}(H'_{ii.1}\Omega^0 H_{jj.1} \otimes H'_{ii}\Omega^{0\prime}M_x\Omega^0 H_{jj})S_{q_{0j}}\|.$$

Although this result does not seem to be much of a simplification over the general result of equation <4.4.2.2>, a great simplification does occur in the limiting distribution of certain characteristic roots of $R_n^*(\hat{\pi})$, which will be established in a later subsection.

4.4.4. The distribution problem remaining to be solved is how to find the limit distribution of the characteristic roots of $R_n(\tilde{\pi})$ and $R_n^*(\hat{\pi})$, given knowledge of the limit distribution of the distinct elements of these matrices. The paper by Anderson [1963] provides a method for finding this distribution: unfortunately, however, it requires minor generalisation to cope with the nature of these limit distributions, and for the possibility of multiple zero roots of the population criterion matrices $R_n(\pi_0)$ and $R_n^*(\pi_0)$. This latter generalisation is required in order to be able to discuss "sequential tests of rank", and also to be able to construct a "test of consistency" - that is, a test of the demand that there exists a solution to equation <4.1.1.1>.

4.5. The Limiting Distribution of the Characteristic Roots of Random Symmetric Matrices

4.5.1. It is perhaps worth stating clearly at the outset that even a sketch of the argument given by Anderson [1963] for finding this limit distribution is rather involved; none the less, this is attempted in the current subsection, proofs being relegated to the succeeding subsection.

Consider the following problem: given three r x r positive semi-definite symmetric matrices $\hat{A}_n$, A_n and A, $\hat{A}_n$ is an estimator of the non-stochastic matrix A_n, which in turn is supposed to be such that

$A_n \rightarrow A.$

In addition, it is assumed that

$\hat{A}_n \xrightarrow{P} A,$

and

$$n^{\frac{1}{2}}v(\hat{A}_n - A_n) \overset{a}{\sim} N(0,\Psi[v(\hat{A}_n);v(A)]). \qquad \langle 4.5.1.1\rangle$$

The matrices A_n and A are assumed to have characteristic roots with an identical multiple root structure:

$A_n = \Gamma_n \Upsilon_n \Gamma_n', \quad A = \Gamma \Upsilon \Gamma',$

where

$$\Upsilon_n = \underset{i=1}{\overset{t}{\oplus}} \upsilon_{in} I_{h_i}, \quad \Upsilon = \underset{i=1}{\overset{t}{\oplus}} \upsilon_i I_{h_i}, \qquad \langle 4.5.1.2\rangle$$

so that h_i is the multiplicity of the distinct characteristic roots υ_{in}, υ_i, $i = 1,\ldots,t$. The distinct characteristic roots are arranged in decreasing order:

$\upsilon_{1n} \geqq \ldots \geqq \upsilon_{tn}; \quad \upsilon_1 \geqq \ldots \geqq \upsilon_t,$

whilst the smallest roots are zero:

$\upsilon_{tn} = \upsilon_t = 0.$

Denote the characteristic roots of the sample matrix $\hat{A}_n$ by $\hat{\upsilon}_{in}$, $i = 1,\ldots,r$: the ordered roots

$\hat{\upsilon}_{1n} \geqq \ldots \geqq \hat{\upsilon}_{rn}$

can be thought of as providing (possibly multiple) estimates of the population roots $\upsilon_{1n},\ldots,\upsilon_{tn}$, so that some notation for associating the h_i "ith largest roots" with the root υ_{in} is helpful. Let

$$\mathbf{p}(j) = \sum_{i=1}^{j} h_i, \quad j = 1,\ldots,\mathbf{t} \qquad \langle 4.5.1.3\rangle$$

so that

$\mathbf{p}(\mathbf{t}) = r,$

and partition up the diagonal matrix of characteristic roots of $\hat{A}_n$, $\hat{\Upsilon}_n$, into $\mathbf{t}$ diagonal blocks of dimension $h_i \times h_i$:

$\hat{\Upsilon}_n = \bigoplus_{j=1}^{\mathbf{t}} \hat{\Upsilon}_{jn};$

then,

$\hat{\Upsilon}_{jn} = \mathrm{diag}\{\hat{\upsilon}_{\mathbf{p}(j-1)+1,n},\ldots,\hat{\upsilon}_{\mathbf{p}(j),n}\}.$

One can then also write

$\hat{\Upsilon}_n - \Upsilon_n = \bigoplus_{i=1}^{\mathbf{t}}(\hat{\Upsilon}_{in} - \upsilon_{in}I_{h_i}).$

The proof strategy adopted concentrates on finding the (marginal) limiting distribution of the diagonal elements of

$n^{1/2}(\hat{\Upsilon}_{in} - \upsilon_{in}I_{h_i});$

that is, the joint distribution of

$n^{1/2}(\hat{\upsilon}_{\mathbf{p}(i-1)+1,n} - \upsilon_{in},\ldots,\hat{\upsilon}_{\mathbf{p}(i),n} - \upsilon_{in})$

for each $i = 1,\ldots,\mathbf{t}$: note that $\mathbf{p}(0)$ is defined to be zero.

Let $\mathcal{U}_n$ be the matrix

$$\mathcal{U}_n = \Gamma_n'[n^{1/2}(\hat{A}_n - A_n)]\Gamma_n$$
$$= n^{1/2}[\Gamma_n'\hat{A}_n\Gamma_n - \Upsilon_n], \qquad \langle 4.5.1.4\rangle$$

so that

$$v(\mathcal{U}_n) \overset{a}{\sim} N[0,\Gamma'\Psi(v(\hat{A}_n);v(A))\Gamma]. \qquad \langle 4.5.1.5\rangle$$

A specific expression for $\mathcal{U}_n$, to be derived, is critical for the proof: to obtain this, note that the matrix

$$\Gamma_n'\hat{A}_n\Gamma_n$$

has the same characteristic roots as $\hat{A}_n$, by the orthogonality of Γ_n. Let the canonical form of this matrix be denoted

$$\Gamma_n'\hat{A}_n\Gamma_n = \mathcal{E}_n\hat{\Upsilon}_n\mathcal{E}_n': \qquad \langle 4.5.1.6\rangle$$

then, by regarding equation <4.5.1.4> as an equation in

$$\Gamma_n'\hat{A}_n\Gamma_n,$$

one can write

$$\Gamma_n'\hat{A}_n\Gamma_n = \Upsilon_n + n^{-1/2}\mathcal{U}_n. \qquad \langle 4.5.1.7\rangle$$

Finally, define the diagonal matrix

$$\mathcal{W}_n = n^{1/2}(\hat{\Upsilon}_n - \Upsilon_n),$$

so that

$$\hat{\Upsilon}_n = \Upsilon_n + n^{-1/2}\mathcal{W}_n.$$

Then one has both

$$\Gamma_n'\hat{A}_n\Gamma_n = \mathcal{E}_n(\Upsilon_n + n^{-1/2}\mathcal{W}_n)\mathcal{E}_n'$$

and

$$\Gamma_n'\hat{A}_n\Gamma_n = \Upsilon_n + n^{-1/2}\mathcal{U}_n.$$

Equating these two expressions, one obtains

$$\mathcal{U}_n = n^{1/2}\{\mathcal{E}_n(\Upsilon_n + n^{-1/2}\mathcal{W}_n)\mathcal{E}_n' - \Upsilon_n\}: \qquad \langle 4.5.1.8\rangle$$

this is the desired expression.

The second stage of the argument then requires the partitioning of the matrices $\mathcal{E}_n$, $\mathcal{U}_n$ and the diagonal matrix $\mathcal{W}_n$ by rows and columns according to the multiple root structure of the problem: let

$$\mathcal{E}_n = \|\mathcal{E}_{ijn}\|, \qquad \mathcal{U}_n = \|\mathcal{U}_{ijn}\|,$$

each block being $h_i \times h_j$,

$$\mathcal{W}_n = \underset{i=1}{\overset{t}{\oplus}} \mathcal{W}_{in},$$

with

$$\mathcal{W}_{in} = n^{1/2}(\hat{\Upsilon}_{in} - \nu_{in} I_{h_i}); \qquad \langle 4.5.1.9 \rangle$$

it will also be helpful to define

$$\mathcal{F}_{ijn} = n^{1/2}\mathcal{E}_{ijn}, \qquad i \neq j,\ i,j = 1,\ldots,\mathbf{t}.$$

It is then necessary to multiply out equation $\langle 4.5.1.8 \rangle$ in partitioned form, along with the orthogonality condition for the characteristic vectors $\mathcal{E}_n$:

$$\mathcal{E}_n\mathcal{E}_n' = I_r.$$

By using the latter equations, one can obtain the following expressions:

$$\mathcal{U}_{iin} = \mathcal{E}_{iin}\mathcal{W}_{in}\mathcal{E}_{iin}' + n^{-1/2}(\mathcal{M}_{iin} - \nu_{in}\mathcal{L}_{iin}) + n^{-1}\mathcal{Y}_{iin}, \qquad i = 1,\ldots,\mathbf{t}-1; \qquad \langle 4.5.1.10 \rangle$$

$$\mathcal{U}_{\mathbf{ttn}} = \mathcal{E}_{\mathbf{ttn}}\mathcal{W}_{\mathbf{tn}}\mathcal{E}_{\mathbf{ttn}}' + n^{-1/2}\mathcal{M}_{\mathbf{ttn}} + n^{-1}\mathcal{Y}_{\mathbf{ttn}}; \qquad \langle 4.5.1.11 \rangle$$

$$\mathcal{U}_{ijn} = \nu_{in}\mathcal{E}_{iin}\mathcal{F}_{jin}' + \nu_{jn}\mathcal{F}_{ijn}\mathcal{E}_{jjn}' + n^{-1/2}\mathcal{M}_{ijn} + n^{-1}\mathcal{Y}_{ijn}, \qquad i,j = 1,\ldots,\mathbf{t}-1,\ i \neq j; \qquad \langle 4.5.1.12 \rangle$$

$$\mathcal{U}_{i\mathbf{t}n} = \nu_{in}\mathcal{E}_{iin}\mathcal{F}_{\mathbf{t}in}' + n^{-1/2}\mathcal{M}_{i\mathbf{t}n} + n^{-1}\mathcal{Y}_{i\mathbf{t}n} \qquad i \neq \mathbf{t}; \qquad \langle 4.5.1.13 \rangle$$

$$\mathcal{U}_{\mathbf{t}jn} = \nu_{jn}\mathcal{F}_{\mathbf{t}jn}\mathcal{E}_{jjn}' + n^{-1/2}\mathcal{M}_{\mathbf{t}jn} + n^{-1}\mathcal{Y}_{\mathbf{t}jn} \qquad j \neq \mathbf{t}; \qquad \langle 4.5.1.14 \rangle$$

$$0 = \mathcal{E}_{iin}\mathcal{F}_{jin}' + \mathcal{F}_{ijn}\mathcal{E}_{jjn}' + n^{-1/2}\mathcal{L}_{ijn}, \qquad i \neq j. \qquad \langle 4.5.1.15 \rangle$$

The matrices $\mathcal{L}_{ijn}$, $\mathcal{M}_{ijn}$, $\mathcal{Y}_{ijn}$ in these expressions are defined in subsection 4.A.1. in the Appendix to this Chapter.

In this collection of equations, it is known, from equation <4.5.1.5>, that

$\mathcal{U}_n \xrightarrow{d} \mathcal{U}$;

the elements of the matrices $\mathcal{E}_n$ and $\mathcal{V}_n$ are functions of $\mathcal{U}_n$, but this function itself depends on n, from the effects of the terms in $n^{-1/2}$ and n^{-1} in equations <4.5.1.10>- <4.5.1.15>. Attempting to show that the random matrices $\mathcal{E}_n$ and $\mathcal{V}_n$ converge in distribution to limiting random matrices $\mathcal{E}$ and $\mathcal{V}$, say, depends on the application of "Rubin's Theorem", which is given by Anderson [1963], and in an apparently different guise by Billingsley [1968, Theorem 5.5].

This theorem may be expressed in the following way. Let $\mathbf{x}_n$ be a random vector converging in distribution to a random vector $\mathbf{x}$:

$\mathbf{x}_n \xrightarrow{d} \mathbf{x}$,

and let $\{\mathbf{f}_n(.)\}$ be a sequence of functions such that

$\mathbf{f}_n(\mathbf{x}_n) \longrightarrow \mathbf{f}(\mathbf{x})$,

for every continuity point of $\mathbf{f}(.)$, when

$\mathbf{x}_n \longrightarrow \mathbf{x}$.

If the probability of the set of discontinuities of the function $\mathbf{f}(.)$ according to the distribution of $\mathbf{f}(\mathbf{x})$ is zero, then

$\mathbf{f}_n(\mathbf{x}_n) \xrightarrow{d} \mathbf{f}(\mathbf{x})$.

The application of this theorem is to show that if one regards $\mathbf{f}_n(\mathbf{x}_n)$ as being defined by equations <4.5.1.10> - <4.5.1.15>, then there exists a similar system of equations, given below, playing the role of $\mathbf{f}(\mathbf{x})$. In writing out these "limiting equations", it is helpful to recall that

$\|\mathcal{U}_{ijn}\| = \mathcal{U}_n \xrightarrow{d} \mathcal{U} = \|\mathcal{U}_{ij}\|$,

and

$v_{in} \rightarrow v_i, \quad i = 1,\ldots,t$

has been assumed. The theorem establishes the existence of the random matrices $\mathcal{E}$ and $\mathcal{W}$ such that

$\|\mathcal{E}_{ijn}\| = \mathcal{E}_n \xrightarrow{d} \mathcal{E}$,

where the diagonal blocks of $\mathcal{E}$ are $\mathcal{E}_{ii}$, $i = 1,\ldots,t$, the off-diagonal blocks are $\mathcal{F}_{ij}$, $i,j = 1,\ldots,t$, $i \neq j$, and

$\mathcal{W}_n = \bigoplus_{i=1}^{t}\mathcal{W}_{in} \xrightarrow{d} \mathcal{W} = \bigoplus_{i=1}^{t}\mathcal{W}_i$.

The proof of these statements is given in the next subsection.

The limiting equations are

$$\mathcal{U}_{ii} = \mathcal{E}_{ii}\mathcal{W}_i\mathcal{E}'_{ii}, \quad i = 1,\ldots,t, \qquad \langle 4.5.1.16\rangle$$

where $\mathcal{E}_{ii}$ is an orthogonal matrix:

$$\mathcal{E}_{ii}\mathcal{E}'_{ii} = I_{h_i}; \qquad \langle 4.5.1.17\rangle$$

$$\mathcal{U}_{ij} = v_i\mathcal{E}_{ii}\mathcal{F}'_{ji} + v_j\mathcal{F}_{ij}\mathcal{E}'_{jj}, \quad i,j = 1,\ldots,t \qquad \langle 4.5.1.18\rangle$$

(with $v_t = 0$);

$$0 = \mathcal{E}_{ii}\mathcal{F}'_{ji} + \mathcal{F}_{ij}\mathcal{E}'_{jj}, \quad i \neq j. \qquad \langle 4.5.1.19\rangle$$

The implications of these equations are not immediately obvious, but it can be seen that the distribution of each $\mathcal{W}_i$ is the limiting distribution of the characteristic roots of the random symmetric matrix $\mathcal{U}_{ii}$. Typically, such distributions depend, in a complex way, on the characteristic vectors $\mathcal{E}_{ii}$ as well - see, for example, Srivastava and Khatri [1979,pp276-291], for the case where $\mathcal{U}_{ii}$ has a central Wishart distribution: this is not the case in the intended application.

One can see, however, that

$\operatorname{tr} U_{ii} = \operatorname{tr} W_i, \quad i = 1,\ldots,t;$

it then follows that

$\operatorname{tr} W_{in} \overset{a}{\approx} \operatorname{tr} U_{ii}.$

That is, the sum of a certain number of the characteristic roots of $\hat{A}_n$ corresponding to the population root v_{in} has a limit distribution given by the limit normal distribution of $\operatorname{tr} U_{ii}$ obtained from equation <4.5.1.5>. It is clear from this equation that the limit distribution depends on the population characteristic vectors in the matrix Γ in a fairly complex way. However, it will be shown that consistent estimation of the appropriate function of these characteristic vectors is possible.

Since the diagonal matrix W_{in} is the diagonal matrix of sampling errors

$n^{1/2}(\hat{v}_{jn} - v_{in})$

(where $j = \mathbf{p}(i-1)+1,\ldots,\mathbf{p}(i)$, $i = 1,\ldots,\mathbf{t}$), these results will provide a limit normal distribution for

$\operatorname{tr} W_{in} = n^{1/2}\sum(\hat{v}_{jn} - v_{in}),$

where the sum ranges over the values of j stated above. A consistent estimator of the variance of this normal distribution can also be found.

It seems appropriate now to fill in the proof details, before discussing the nature of this limit normal distribution and its application.

4.5.2 In the application of Rubin's Theorem, it will be necessary to use the Sturmian Separation Theorem for the characteristic roots of a real symmetric matrix (see Bellman [1970, p117]): let $\mathbf{A}$ be an $\mathbf{n} \times \mathbf{n}$ real symmetric matrix whose $i \times i$ principal submatrices are denoted $\mathbf{A}_i$, $i = 1,\ldots,\mathbf{n}$, and let $ch_j(\mathbf{A})$ denote the jth largest root of $\mathbf{A}$. Then,

$$ch_j(\mathbf{A}_{i+1}) \geq ch_j(\mathbf{A}_i) \geq ch_{j+1}(\mathbf{A}_{i+1}), \quad i,j = 1,\ldots,\mathbf{n},\ j \leq i. \tag{4.5.2.1}$$

The proof strategy is to show that as

$$\mathcal{U}_n \longrightarrow \mathcal{U},$$

in the non-stochastic sense required to verify the conditions of Rubin's Theorem, then

$$\mathcal{W}_{in} \longrightarrow \mathcal{W}_i;$$

when this has been established for each $i = 1,\ldots,\mathbf{t}$, the convergences

$$\mathcal{E}_{iin} \longrightarrow \mathcal{E}_{ii}, \quad i = 1,\ldots,\mathbf{t},$$

$$\mathcal{F}_{ijn} \longrightarrow \mathcal{F}_{ij}, \quad i,j = 1,\ldots,\mathbf{t},\ i \neq j,$$

are deduced. The essence of the proof to be given below is given by Anderson [1963], although some of the details differ.

Consider first the diagonal elements of $\mathcal{W}_{1n}$:

$$\mathcal{W}_{1n} = \mathrm{diag}\{n^{½}(\hat{\upsilon}_{1n} - \upsilon_{1n}),\ldots,n^{½}(\hat{\upsilon}_{\mathbf{p}(1),n} - \upsilon_{1n})\};$$

here, $\hat{\upsilon}_{1n},\ldots,\hat{\upsilon}_{\mathbf{p}(1),n}$ are the $\mathbf{p}(1)$ largest roots of $\hat{\mathcal{A}}_n$ and hence of

$$\Gamma_n'\hat{\mathcal{A}}_n\Gamma_n = \Upsilon_n + n^{-½}\mathcal{U}_n,$$

by equation <4.5.1.7>. It then follows that the diagonal elements of $\mathcal{V}_{1n}$ are the $\mathbf{p}(1)$ largest roots of the matrix

$$\Upsilon_n + n^{-\frac{1}{2}}\mathcal{U}_n - \nu_{1n}\mathbf{I}_r: \tag{4.5.2.2}$$

that is, of the determinental equation (in the variable υ)

$$0 = \det[\Upsilon_n + n^{-\frac{1}{2}}\mathcal{U}_n - (\nu_{1n} + \upsilon)\mathbf{I}_r], \tag{4.5.2.3}$$

since the roots of the matrix in equation <4.5.2.2> are $\hat{\upsilon}_{in} - \nu_{1n}$, $\quad i = 1,\ldots,r$.

Partition up Υ_n and $\mathcal{U}_n$ into the first $\mathbf{p}_1 = \mathbf{p}(1)$ rows and columns, and the remaining $r - \mathbf{p}_1$ rows and columns, say,

$$\Upsilon_n = \begin{bmatrix} \Upsilon_{1n} & : & 0 \\ 0 & : & \Upsilon_{1n}^* \end{bmatrix} = \begin{bmatrix} \nu_{1n}\mathbf{I}_{\mathbf{p}_1} & : & 0 \\ 0 & : & \Upsilon_{1n}^* \end{bmatrix}$$

$$\mathcal{U}_n = \begin{bmatrix} \mathcal{U}_{11n} & : & \mathcal{U}_{1n}^{*\prime} \\ \mathcal{U}_{1n}^* & : & \mathcal{U}_{11n}^* \end{bmatrix},$$

so that Υ_{1n}^* and $\mathcal{U}_{11n}^*$ are the "complementary" principal submatrices to Υ_{1n} and $\mathcal{U}_{11n}$ respectively. Then, the determinental equation <4.5.2.3> may be written as

$$0 = \det\begin{bmatrix} n^{-\frac{1}{2}}(\mathcal{U}_{11n}-\upsilon\mathbf{I}_{\mathbf{p}_1}): & n^{-\frac{1}{2}}\mathcal{U}_{1n}^{*\prime} \\ n^{-\frac{1}{2}}\mathcal{U}_{1n}^* & :\Upsilon_{1n}^*-\nu_{1n}\mathbf{I}_{r-\mathbf{p}_1}+n^{-\frac{1}{2}}[\mathcal{U}_{11n}^*-\upsilon\mathbf{I}_{r-\mathbf{p}_1}] \end{bmatrix}; \tag{4.5.2.4}$$

as a monic polynomial in $n^{-\frac{1}{2}}\upsilon$, this can be written as

$$\mathbf{f}_n(\upsilon) = n^{-\frac{1}{2}\mathbf{t}}\prod_{i=1}^{\mathbf{t}}(\hat{\upsilon}_{in} - \nu_{1n} - \upsilon).$$

Consider now the effect of factoring out $n^{-\frac{1}{2}}$ from the first $\mathbf{p}_1$ columns of the matrix in <4.5.2.4> to produce

$$0 = n^{-\frac{1}{2}\mathbf{p}_1}\det\begin{bmatrix} \mathcal{U}_{11n} - \upsilon I_{h_1} & : & n^{-\frac{1}{2}}\mathcal{U}^{*\prime}_{1n} \\ \mathcal{U}^{*}_{1n} & : & \Upsilon^{*}_{1n} - \nu_{1n}I_{t-h_1} + n^{-\frac{1}{2}}(\mathcal{U}^{*}_{11n} - \upsilon I_{t-h_1}) \end{bmatrix};$$

<4.5.2.5>

this determinant may similarly be written as a monic polynomial in $n^{-\frac{1}{2}}\upsilon$:

$$\mathbf{f}^{*}_{n}(\upsilon) = n^{-\frac{1}{2}(r-\mathbf{p}_1)}\prod_{i=1}^{r}(\hat{\upsilon}_{in} - \nu_{1n} - \upsilon) \qquad \langle 4.5.2.6\rangle$$

$$= (\hat{\upsilon}_{i_1 n} - \nu_{1n} - \upsilon)\ldots(\hat{\upsilon}_{i(\mathbf{p}_1)n} - \nu_{1n} - \upsilon)n^{-\frac{1}{2}(r-\mathbf{p}_1)}$$

$$\times\ \Pi'(\hat{\upsilon}_{jn} - \nu_{1n} - \upsilon)$$

for some selection of $\mathbf{p}_1$ indices, $i_1,\ldots,i(\mathbf{p}_1) = i_{\mathbf{p}_1}$ from $i = 1,\ldots,r$; here, Π' denotes the product over the remaining $r - \mathbf{p}_1$ indices. Which selection of $\mathbf{p}_1$ indices from r are $i_1,\ldots,i_{\mathbf{p}_1}$ will be established shortly.

Under the hypothesised condition $\mathcal{U}_n \rightarrow \mathcal{U}$,

$$n^{-\frac{1}{2}}\mathcal{U}^{*\prime}_{1n} \rightarrow 0,$$

$$n^{-\frac{1}{2}}(\mathcal{U}^{*}_{11n} - \upsilon I_{r-\mathbf{p}_1}) \rightarrow 0,$$

so that

$$\mathbf{f}^{*}_{n}(\upsilon) \rightarrow \det(\mathcal{U}_{11} - \upsilon I_{\mathbf{p}_1})\det(\Upsilon^{*}_{1} - \nu_1 I_{r-\mathbf{p}_1}),$$

$$= \mathbf{f}^{*}(\upsilon),$$

say, where

$$\begin{bmatrix} \Upsilon_{1n} & : & 0 \\ 0 & : & \Upsilon^{*}_{1n} \end{bmatrix} = \Upsilon_n \rightarrow \Upsilon = \begin{bmatrix} \Upsilon_1 & : & 0 \\ 0 & : & \Upsilon^{*}_{1} \end{bmatrix} = \begin{bmatrix} \nu_1 I_{\mathbf{p}_1} & : & 0 \\ 0 & : & \Upsilon^{*}_{1} \end{bmatrix}.$$

<4.5.2.7>

The solutions to

$$\mathbf{f}^{*}_{n}(\upsilon) = 0$$

will approach, by continuity, the solutions to

$$\mathbf{f}^{*}(\upsilon) = 0:$$

the trick is to show that the $\mathbf{p}_1$ largest roots

$\hat{\upsilon}_{in} - \upsilon_{1n}, \quad i = 1,\ldots,\mathbf{p}_1$

approach the roots of

$\det(\mathcal{U}_{11} - \upsilon I_{\mathbf{p}_1}) = 0.$

To establish this, first note that by the Sturmian Separation Theorem, equation <4.5.2.1>,

$$\hat{\upsilon}_{in} - \upsilon_{1n} \geq ch_i(\mathcal{U}_{11n}), \quad i = 1,\ldots,\mathbf{p}_1$$
$$\geq 0,$$

since $\mathcal{U}_{11n}$ is a positive semidefinite matrix. Thus,

$$\lim_{n\to\infty}(\hat{\upsilon}_{in} - \upsilon_{1n}) \geq \lim_{n\to\infty} ch_i(\mathcal{U}_{11n}) = ch_i(\mathcal{U}_{11}) \geq 0,$$
$$i = 1,\ldots,\mathbf{p}_1. \qquad \text{<4.5.2.8>}$$

Next, note that from equation <4.5.2.6> and the convergence of $\mathbf{f}_n^*(\upsilon)$ to $\mathbf{f}^*(\upsilon)$ that for some indices $j = 1,\ldots,r$ and $k = 2,\ldots,\mathbf{t}$,

$$n^{-\frac{1}{2}}(\hat{\upsilon}_{jn} - \upsilon_{1n} - \upsilon) \longrightarrow \upsilon_k - \upsilon_1 < 0.$$

This in turn implies that the function of υ defined by

$$n^{-\frac{1}{2}}(\hat{\upsilon}_{jn} - \upsilon_{1n}) = n^{-\frac{1}{2}}\upsilon + \upsilon_k - \upsilon_1$$

converges to

$$\upsilon_k - \upsilon_1 < 0,$$

pointwise in υ.

However, from equation <4.5.2.8> above,

$$n^{-\frac{1}{2}}(\hat{\upsilon}_{jn} - \upsilon_{1n}) \longrightarrow 0, \ j = 1,\ldots,\mathbf{p}_1,$$

so that for $j = \mathbf{p}_1+1,\ldots,r,$ it must be true that

$$n^{-\frac{1}{2}}(\hat{\upsilon}_{jn} - \upsilon_{1n}) \longrightarrow \upsilon_k - \upsilon_1 < 0$$

for some $k = 2,\ldots,\mathbf{t}$. The roots of <4.5.2.5> have therefore been split into two subsets, of $j = 1,\ldots,\mathbf{p}_1$, and $j = \mathbf{p}_1+1,\ldots,\mathbf{t}$, the latter subset converging to a negative number, and the former converging to the roots of $\mathcal{U}_{ii}$; this argument is making use of the continuity of the roots of the polynomial $f_n^*(v)$ in its coefficients.

Thus, it has been shown that the p_1 largest roots of <4.5.2.4>, contained in the diagonal matrix W_{in}, converge to the characteristic roots of $\mathcal{U}_{11}$ contained in the diagonal matrix W_1:

$$W_{1n} \rightarrow W_1.$$

This technique of separating the roots of determinental equations in "partly scaled" matrices like that in equation <4.5.2.5> can be applied to each group of sample characteristic roots corresponding to each distinct population root v_{in}; the use of the Sturmian separation theorem in isolating the desired group of roots is a little more complex however. In particular, to establish the result for the population zero root

$$v_{tn} = v_t = 0,$$

one can reorder rows and columns so that the blocks of $\mathcal{U}_n$ and Υ_n in <4.5.2.2> corresponding to these zero roots are in the 1-1 position: i.e.

$$\begin{bmatrix} \mathcal{U}_{ttn} & : & \mathcal{U}_{tn}^{*\prime} \\ \mathcal{U}_{tn}^* & : & \mathcal{U}_{ttn}^* \end{bmatrix}; \quad \begin{bmatrix} 0 & : & 0 \\ 0 & : & \Upsilon_{tn}^* \end{bmatrix}.$$

The determinant corresponding to <4.5.2.4> is

$$0 = \det\begin{bmatrix} n^{-1/2}(\mathcal{U}_{ttn} - vI_{h_t}) & : & n^{-1/2}\mathcal{U}_{tn}^{*\prime} \\ n^{-1/2}\mathcal{U}_{tn}^* & : & \Upsilon_{tn}^* + n^{-1/2}(\mathcal{U}_{ttn}^* - vI_{h_t}^*) \end{bmatrix},$$

and the diagonal elements of W_{tn} are then the h_t smallest roots of this determinental equation; $I^*_{h_t}$ is the matrix complementary to I_{h_t} in I_r. Factoring out $n^{-1/2}$ from the first h_t rows and columns, and passing to a limit produces

$$0 = \det(\mathcal{U}_{tt} - \upsilon I_{h_t})\det\Upsilon^*_t,$$

where Υ^*_t is the matrix complementary to Υ_t in Υ.

In this case, the Sturmian Separation Theorem <4.5.2.1> then shows that

$$ch_i(\mathcal{U}_{ttn}) \geq \hat{\upsilon}_{p(t-1)+i,n}, \qquad i = 1,\ldots,h_t,$$

so that

$$n^{-1/2}\hat{\upsilon}_{p(t-1)+i,n} \rightarrow 0,$$

whilst

$$n^{-1/2}\hat{\upsilon}_{jn} \rightarrow \upsilon_k > 0$$

for $j = 1,\ldots,p(t-1)$ and some $k = 1,\ldots,t-1$. Thus, by the same arguments as before,

$$\mathcal{W}_{tn} \rightarrow \mathcal{W}_t$$

as $\mathcal{U}_n \rightarrow \mathcal{U}$.

One can now turn to deal with the convergence of the characteristic vectors, in particular, the characteristic vector corresponding to the first characteristic root,

$$\hat{\upsilon}_{1n} - \upsilon_{1n}.$$

This will require a partition of the first h_1 columns of the matrix in <4.5.2.5>, where υ is replaced by $\hat{\upsilon}_{1n}-\upsilon_{1n}$, into 1 and h_1-1 columns, using the following partition of $\mathcal{U}_n$ and $\mathcal{U}^*_{1n}$:

$$\mathcal{U}_{11n} = \begin{bmatrix} u_{11n} & : & u^{*\prime}_{1n} \\ u^*_{1n} & : & \bar{\mathcal{U}}_{11n} \end{bmatrix}, \qquad \mathcal{U}^*_{1n} = [\ \bar{u}^*_{1n} : \bar{\mathcal{U}}^*_{1n}\ ].$$

This yields

$$\left[\begin{array}{c:c:c} u_{11n}-(\hat{v}_{1n}-v_{1n}) & u_{1n}^{*\prime} & \\ u_{1n}^{*} & \mathcal{V}_{11n}-(\hat{v}_{1n}-v_{1n})I_{h_1-1} & n^{-\frac{1}{2}}\bar{u}_{1n}^{*\prime} \\ \bar{u}_{1n}^{*} & \bar{\mathcal{V}}_{1n}^{*} & n^{-\frac{1}{2}}\bar{\mathcal{V}}_{1n}^{*\prime} \\ & & \Upsilon_{1n}^{*}-v_{1n}I_{t-h_1}+n^{-\frac{1}{2}}(\mathcal{U}_{11n}^{*}-(\hat{v}_{1n}-v_{1n})I_{t-h_1}) \end{array}\right]. \quad \langle 4.5.2.9\rangle$$

The elements of the unnormalised characteristic vector corresponding to $\hat{v}_{1n} - v_{1n}$ are easily shown to be proportional to the cofactors of the first column of the matrix above. For the first element, the cofactor is

$$\det\left[\begin{array}{c:c} \mathcal{V}_{11n} - (\hat{v}_{1n} - v_{1n})I_{h_1-1} & n^{-\frac{1}{2}}\bar{\mathcal{V}}_{1n}^{*\prime} \\ \bar{\mathcal{V}}_{1n}^{*} & \Upsilon_{1n}^{*} - v_{1n}I_{t-h_1} + n^{-\frac{1}{2}}(\mathcal{U}_{11n}^{*} - (\hat{v}_{1n} - v_{1n})I_{t-h_1}) \end{array}\right]:$$

using a Laplace expansion based on the minors of the first h_1-1 rows, it can be expressed as

$\mathbf{a}_{1n} + n^{-\frac{1}{2}}\mathbf{b}_{1n}$,

$\mathbf{b}_{1n}$ containing some terms in powers of $n^{-\frac{1}{2}}$, but such that

$\mathbf{a}_{1n} \longrightarrow \mathbf{a}_1, \quad \mathbf{b}_{1n} \longrightarrow \mathbf{b}_1,$

because $\mathcal{U}_n \longrightarrow \mathcal{U}$, $\Upsilon_n \longrightarrow \Upsilon$, $W_n \longrightarrow W$ is assumed. Similarly, one can show that the other cofactors of the first column of ⟨4.5.2.9⟩ have the same form,

$\mathbf{a}_{in} + n^{-\frac{1}{2}}\mathbf{b}_{in}, \quad i \leq h_1,$

but the form

$n^{-\frac{1}{2}}\mathbf{b}_{in}, \quad i = h_1+1,\ldots,t,$

where

$\mathbf{a}_{in} \longrightarrow \mathbf{a}_i, \quad \mathbf{b}_{in} \longrightarrow \mathbf{b}_i, \quad i = 1,\ldots,t.$

Following normalisation, one can conclude that the first h_1 elements of the "first" characteristic vector converges, and $n^{1/2}$ times the remaining $t-h_1$ elements converges. In turn, this argument can be applied to the characteristic vectors associated with each of the diagonal elements of the matrix W_{1n}.

Hence, as $U_n \rightarrow U$,

$E_{iin} \rightarrow E_{ii}$, $\quad i = 1,\ldots,t$,

$F_{ijn} = n^{1/2}E_{ijn} \rightarrow F_{ij}$, $\; i,j = 1,\ldots,t$, $\; i \neq j$.

To establish that the limiting equations

<4.5.1.16>-<4.5.1.19>,

$U_{ii} = E_{ii}W_iE'_{ii}$, $\quad i = 1,\ldots t$,

$E_{ii}E'_{ii} = I_{h_i}$,

$U_{ij} = v_iE_{ii}F'_{ji} + v_jF_{ij}E'_{jj}$, $\quad i,j = 1,\ldots,t$,

$0 = E_{ii}F'_{ji} + F_{ij}E'_{jj}$, $\quad i,j = 1,\ldots,t$, $\; i \neq j$

are valid, one can return to the partitioned equations <4.5.1.10>- <4.5.1.15> and show that as $U_n \rightarrow U$, the terms

$n^{-1/2}M_{ijn}$, $n^{-1/2}L_{ijn}$, $n^{-1}Y_{ijn}$

for $i, j = 1,\ldots,t$ all converge to zero, and the remaining terms converge to their limits, reproducing the equations <4.5.1.16>-<4.5.1.19> above. Since it has been assumed that the matrix $\hat{A}_n$ has distinct roots everywhere, the derived matrix U_n also has distinct roots: according to Anderson [1963], this is the only possible source of discontinuities in the limit function $f(x)$ of Rubin's Theorem.

<u>4.5.3.</u> All the conditions required by Rubin's Theorem

having been met, it follows that for each $i = 1,\ldots,t$, the limiting distribution of the diagonal elements of

$$\hat{\Upsilon}_n - \Upsilon_n = \mathop{\oplus}_{i=1}^{t}(\hat{\Upsilon}_{in} - \Upsilon_{in}) = \mathop{\oplus}_{i=1}^{t}W_{in}$$

(see equation <4.5.1.9>) is determined from the limiting equations <4.5.1.16>-<4.5.1.19>, repeated at the end of the preceding subsection. Specifically,

$$\mathrm{tr}\ W_{iin} \overset{a}{\sim} \mathrm{tr}\ U_{iin} \overset{a}{\sim} \mathrm{tr}\ U_{ii} = \mathrm{tr}\ W_i.$$

Recall from equations <4.5.1.4> and <4.5.1.1> that

$$U_n = \Gamma_n'[n^{1/2}(\hat{A}_n - A_n)]\Gamma_n,$$

where Γ_n is the matrix of characteristic vectors of the population matrix

$$A_n = \Gamma_n\Upsilon_n\Gamma_n',$$

and

$$n^{1/2}\mathrm{v}(\hat{A}_n - A_n) \overset{a}{\sim} N(0,\Psi[\mathrm{v}(\hat{A}_n);\mathrm{v}(A)]).$$

By partitioning up the matrix Γ_n by columns to match the multiple root structure of A_n described in equation <4.5.1.2>,

$$\Gamma_n = [\Gamma_{1n},\ldots,\Gamma_{tn}], \qquad \text{<4.5.3.1>}$$

with corresponding population matrix

$$\Gamma = [\Gamma_1,\ldots,\Gamma_t],$$

it can be seen that

$$\begin{aligned}\mathrm{tr}\ U_{iin} &= \mathrm{tr}[\Gamma_{in}'n^{1/2}(\hat{A}_n - A_n)\Gamma_{in}] \\ &= \mathrm{tr}[n^{1/2}(\hat{A}_n - A_n)\Gamma_{in}\Gamma_{in}'] \\ &= (\mathrm{vec}\ \Gamma_{in}\Gamma_{in}')'\mathrm{vec}\ n^{1/2}(\hat{A}_n - A_n) \\ &= (\mathrm{vec}\ I_{h_i})'(\Gamma_{in}' \otimes \Gamma_{in}')\mathrm{vec}\ n^{1/2}(\hat{A}_n - A_n) \\ &= (\mathrm{vec}\ I_{h_i})'(\Gamma_{in}' \otimes \Gamma_{in}')\mathbf{D}_r'n^{1/2}\mathrm{v}(\hat{A}_n - A_n) \qquad \text{<4.5.3.2>}\end{aligned}$$

(using equations <1.6.1.2> and <1.6.3.10>) has a limit normal distribution with mean zero and a variance that can be obtained fairly easily once

$\Psi(v(\hat{A}_n);v(A))$

is known.

Thus, a limit normal distribution for sums of the sampling errors of the characteristic roots of $\hat{A}_n$ corresponding to the distinct population roots has been obtained; these distributions will form the basis of tests of the rank and consistency conditions for identification - see equation <4.1.1.2>. To make these limiting distributions operational, it is necessary to find a consistent estimator of $\Gamma_{in}\Gamma'_{in}$, or more accurately, of $\Gamma_i\Gamma'_i$. Let the canonical form of $\hat{A}_n$ be

$$\hat{A}_n = K_n\hat{\Upsilon}_nK'_n.$$

Recall from equation <4.5.1.6> that

$$\Gamma'_n\hat{A}_n\Gamma_n = \mathcal{E}_n\hat{\Upsilon}_n\mathcal{E}'_n,$$

so that

$$\Gamma'_nK_n = \mathcal{E}_n,$$

or

$$K_n = \Gamma_n\mathcal{E}_n:$$

K_n can be regarded as an estimator of Γ_n, or its limit Γ. Now, partition up K_n to match the partitioning of Γ_n and Γ, and similarly for $\mathcal{E}_n$:

$$K_n = [K_{1n},\ldots,K_{tn}],$$

$$\mathcal{E}_n = [\mathcal{E}_{1n},\ldots,\mathcal{E}_{tn}],$$

where

$$\mathcal{E}'_{in} = [\mathcal{E}'_{1in},\ldots,\mathcal{E}'_{tin}],$$

using the previous partition of $\mathcal{E}_n$. Then,

$$K_{in} = \Gamma_n\mathcal{E}_{in},$$

and the estimator of $\Gamma_{in}\Gamma'_{in}$ (or $\Gamma_i\Gamma'_i$) can be written as

$$K_{in}K'_{in} = \Gamma_n\mathcal{E}_{in}\mathcal{E}'_{in}\Gamma'_n.$$

The matrix $\mathcal{E}_{in}\mathcal{E}'_{in}$ has the form

$$\mathcal{E}_{in}\mathcal{E}'_{in} = \|\mathcal{E}_{jin}\mathcal{E}'_{kin}\|, \quad j,k = 1,\ldots,t:$$

it was shown in subsection 4.5.2. that

$$\mathcal{E}_{iin} \xrightarrow{d} \mathcal{E}_{ii}$$

such that

$$\mathcal{E}_{ii}\mathcal{E}'_{ii} = I_{h_i},$$

and

$$F_{ijn} = n^{1/2}\mathcal{E}_{ijn} \xrightarrow{d} F_{ij},$$

so that

$$\mathcal{E}_{ijn} \xrightarrow{d} 0, \quad i \neq j.$$

Thus,

$$\mathcal{E}_{in}\mathcal{E}'_{in} \xrightarrow{d} \mathcal{D},$$

say, a constant block matrix, with a unit matrix of dimension h_i in the ith row and column block, all other blocks being null matrices.

As a result of this,

$$K_{in}K'_{in} \xrightarrow{p} \Gamma\mathcal{D}\Gamma' = \Gamma_i\Gamma'_i,$$

as required: the "in probability" assertion follows from the fact that convergence in distribution to a degenerate random variable is equivalent to convergence in probability to the appropriate constant (see, for example, Serfling [1980, p19]).

4.5.4. These general results will be used simply to justify the existence of a limit normal distribution for the characteristic roots of the rank criterion matrices $R_n(\tilde{\pi}_0)$ and

$R_n^*(\hat{\pi})$ defined in equations <4.4.1.3> and <4.4.1.4>, which may be interpreted as the matrix $\hat{A}_n$ in the general case, the analogous population matrices being $R_n(\pi_0)$ and $R_n^*(\pi_0)$ of equations <4.4.1.1> and <4.4.1.2>. Most of the discussion of tests of rank will be conducted under a null hypothesis of full rank, together with an assumption of distinct population roots (in subsection 4.6.2.). However, these assumptions do not eliminate the practical difficulty that the combination of equation <4.5.3.2> with the limit distribution of

$v(R_n(\tilde{\pi}_0) - R_n(\pi^0))$

say, (see equations <4.4.1.10> and <4.4.1.11>) only produces rather complicated expressions for the variances of the limit distributions of the required characteristic roots, and which will not therefore be given explicitly.

Another justification for this omission is that the multiple root structure of $R_n(\pi_0)$ or $R_n^*(\pi_0)$ does not match the free parameter structure of the structural equations: even in the "usual special case", where $R_n^*(\pi_0)$ becomes block-diagonal, its canonical form is not necessarily block-diagonal, since individual characteristic roots coming from different diagonal blocks of $R_n^*(\pi_0)$ have to be permuted to appear together in the matrix Υ_n.

<u>4.5.5.</u> If, however, one adds to this special case the assumption that the characteristic roots of the population criterion $R_n^*(\pi_0)$ are distinct, a very useful simplification in the limiting distribution of the sample characteristic

roots occurs.

The basic reason for this simplification is that the canonical form of $R_n^*(\pi_0)$ then becomes block diagonal: from equations <4.4.3.1> and <4.4.3.2>,

$R_n^*(\pi_0) = {}_{i=1}^{\oplus m}R_{in}^*(\pi_0) = n^{-1}{}_{i=1}^{\oplus m}H_{ii}'Q_0'X'XQ_0H_{ii}$,

a $q_0 \times q_0$ matrix, and the canonical form of each $R_{in}^*(\pi_0)$ is say

$R_{in}^*(\pi_0) = \Gamma_{iin}^*\Upsilon_{in}^*\Gamma_{iin}^{*\prime};$ <4.5.5.1>

recall that $R_{in}^*(\pi_0)$ is $q_{0i} \times q_{0i}$, $i = 1,\ldots,m$, with

${}_{i=1}^{\sum m}q_{0i} = q_0$.

This involves a slight change of emphasis, as well as notation, in that in the previous analysis,

$\Upsilon_{in} = \nu_{in}I_{h_i}$

with

$\nu_{1n} \geq \ldots \geq \nu_{tn};$

here, any ordering by size is within the matrix Υ_{in}^*, and there is no necessary size relationship between the diagonal elements of Υ_{in}^* and Υ_{jn}^*, $i > j$. In addition,

$t = r = q_0$

and

$h_i = q_{0i}, \quad i = 1,\ldots,m,$

so that

$p_j = {}_{i=1}^{\sum j}q_{0i}, \quad j = 1,\ldots,m$

(with $p_m = q_0$), and finally,

$\Upsilon_{in}^* = \mathrm{diag}\{\nu_{p(i-1)+1,n}^*,\ldots,\nu_{p_i,n}^*\}$.

Let E_i be the matrix defined in equation <1.6.4.1>:

$$E_i = \begin{bmatrix} O_{q_{01},q_{0i}} \\ \cdot \\ \cdot \\ I_{q_{0i}} \\ \cdot \\ \cdot \\ O_{q_{0m},q_{0i}} \end{bmatrix};$$

then, using equation <4.5.3.1>,

$$\Gamma^*_{in} = E_i \Gamma^*_{iin}, \quad i = 1,\ldots,m.$$

In turn, let the columns of Γ^*_{in} and Γ^*_{iin} be

$$\Gamma^*_{in} = [g^*_{i.1n},\ldots,g^*_{i.q_0in}], \quad i = 1,\ldots,m, \qquad \text{<4.5.5.2>}$$

and

$$\Gamma^*_{iin} = [g^*_{ii.1n},\ldots,g^*_{ii.q_0in}]: \quad i = 1,\ldots,m$$

with corresponding limits

$$\Gamma^*_i = [g^*_{i.1},\ldots,g^*_{i.q_0i}], \qquad \text{<4.5.5.3>}$$

$$\Gamma^*_{ii} = [g^*_{ii.1},\ldots,g^*_{ii.q_0i}];$$

then,

$$g^*_{i.jn} = E_i g^*_{ii.jn}, \quad i = 1,\ldots,m, \; j = 1,\ldots,q_{0i}. \qquad \text{<4.5.5.4>}$$

One can now establish that the marginal limiting distribution of the jth smallest root of $R^*_{in}(\hat{\pi})$ (given the assumption of distinct population roots) depends only on the marginal limiting distribution of

$$n^{1/2} v(R^*_{in}(\hat{\pi}) - R^*_{in}(\pi^0)),$$

and not the blocks corresponding to the other equations of the model. For, from the limiting distribution statement in subsection 4.5.3., particularly equation <4.5.3.2>, one can conclude that

$$n^{1/2}(\hat{v}^*_{\mathbf{p}(i-1)+j,n} - v^*_{\mathbf{p}(i-1),n}) \overset{a}{\approx} (g^{*\prime}_{i.jn} \otimes g^{*\prime}_{i.jn}) D'_{q_0} n^{1/2} v(R^*_n(\hat{\pi}) - R^*_n(\pi^0))$$
$$= (g^{*\prime}_{i.jn} \otimes g^{*\prime}_{i.jn}) n^{1/2} \mathrm{vec}\ (R^*_n(\hat{\pi}) - R^*_n(\pi^0))$$
$$= (g^{*\prime}_{ii.jn} E'_i \otimes g^{*\prime}_{ii.jn} E'_i) \overset{m}{\underset{i=1}{\oplus}} (I_{q_{0i}} \otimes E_i)$$
$$\times\ n^{1/2} \begin{bmatrix} \mathrm{vec}(R^*_{1n}(\hat{\pi}) - R^*_{1n}(\pi^0)) \\ \cdot \\ \cdot \\ \mathrm{vec}(R^*_{mn}(\hat{\pi}) - R^*_{mn}(\pi^0)) \end{bmatrix}$$

using equations <4.4.3.4> and <4.5.5.4> above. Now, consider the product

$$(E'_i \otimes E'_i) \overset{m}{\underset{i=1}{\oplus}} (I_{q_{0i}} \otimes E_i) = [0 \ldots 0 : I_{q_{0i}} \otimes E'_i : 0 \ldots 0] \overset{m}{\underset{i=1}{\oplus}} (I_{q_{0i}} \otimes E_i)$$
$$= [0 \ldots 0 : I_{q_{0i}} \otimes E'_i E_i : 0 \ldots 0],$$

and it is easy to see from the definition <1.6.4.2> that $E'_i E_i = I_{q_{0i}}$.

Hence,

$$n^{1/2}(\hat{v}^*_{\mathbf{p}(i-1)+j,n} - v^*_{\mathbf{p}(i-1)+j,n}) \overset{a}{\approx} (g^{*\prime}_{ii.j} \otimes g^{*\prime}_{ii.j})$$
$$\times\ n^{1/2} \mathrm{vec}(R^*_{in}(\hat{\pi}) - R^*_{in}(\pi^0))$$
$$\overset{a}{\approx} N(0, \Psi(\hat{v}^*_{\mathbf{p}(i-1)+j,n};\kappa))$$

(where κ simply represents the relevant parameters) and

$$\Psi(\hat{v}^*_{\mathbf{p}(i-1)+j,n};\kappa) = 4(g^{*\prime}_{ii.j} \otimes g^{*\prime}_{ii.j}) S_{q_{0i}} (H'_{ii.1} \Omega^0 H_{ii.1} \otimes$$
$$H'_{ii} \Omega^{0\prime} M_X \Omega^0 H_{ii}) S_{q_{0i}}$$
$$\times\ (g^*_{ii.j} \otimes g^*_{ii.j}). \qquad \text{<4.5.5.5>}$$

By <1.6.3.8>,

$$S_{q_{0i}} (g^*_{ii.j} \otimes g^*_{ii.j}) = (g^*_{ii.j} \otimes g^*_{ii.j}),$$

and multiplying through the Kronecker product, one obtains

$$\Psi(\hat{v}^*_{\mathbf{p}(i-1)+j,n};\kappa) = 4(g^{*\prime}_{ii.j} H'_{ii.1} \Omega^0 H_{ii.1} g^*_{ii.j} \otimes$$
$$g^{*\prime}_{ii.j} H'_{ii} \Omega^{0\prime} M_X \Omega^0 H_{ii} g^*_{ii.j}).$$

The matrix

$$H_{ii}'\Omega^{0\prime}M_X\Omega^0H_{ii}$$

is actually

$$R_i^*(\pi^0) = \lim_{n\to\infty}R_{in}^*(\pi^0),$$

so that it has canonical form

$$R_i^*(\pi^0) = \Gamma_{ii}^*\Upsilon_i^*\Gamma_{ii}^{*\prime},$$

and it follows that

$$\mathbf{g}_{ii.j}^{*\prime}H_{ii}'\Omega^{0\prime}M_X\Omega^0H_{ii}\mathbf{g}_{ii.j}^* = \upsilon_{\mathbf{p}(i-1)+j}^*.$$

Overall, then,

$$\Psi(\hat{\upsilon}_{\mathbf{p}(i-1)+j,n}^*;\kappa) = 4\upsilon_{\mathbf{p}(i-1)+j}^*\mathbf{g}_{ii.j}^{*\prime}H_{ii.1}' \times \Omega^0H_{ii.1}\mathbf{g}_{ii.j}^*, \qquad \langle 4.5.5.6\rangle$$

which reveals rather clearly the dependence of the limit distribution on the corresponding population characteristic roots. Notice too that the distribution depends only on quantities associated with the ith structural equation, apart from the covariance matrix Ω^0.

4.6. Tests on the Rank of $R_n(\pi_0)$ and $R_n^*(\pi_0)$.

4.6.1. It seems appropriate, after the complex arguments of the preceding two sections, to recall the original purpose. For the null hypothesis simultaneous equations model, the parameter vector γ is identified if and only if there exists a unique solution to the equation

$$(I_m \otimes \Pi_0)(H\gamma + h) = 0$$

for given Π_0. Given the existence of a solution (the "consistency condition"), it is necessary and sufficient for the identification of γ for the matrix

$$(I_m \otimes \Pi_0)H$$

to have full column rank q_0.

The matrices defined in equations <4.4.1.1> and <4.4.1.2>,

$$R_n(\pi_0) = n^{-1}H'(\Sigma_0^{-1} \otimes \Pi_0'X'X\Pi_0)H$$

and

$$R_n^*(\pi_0) = n^{-1}H'(I_m \otimes \Pi_0'X'X\Pi_0)H$$

have the same rank as

$$(I_m \otimes \Pi_0)H,$$

given that X has full column rank k_1, and can be estimated by the matrices of equations <4.4.1.3> and <4.4.1.4> respectively,

$$R_n(\tilde{\pi}_0) = n^{-1}H'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'X\tilde{\Pi}_0)H,$$

which uses the FIML estimates of the null hypothesis model, and

$$R_n^*(\hat{\pi}) = n^{-1}H'(I_m \otimes \hat{\Pi}'X'X\hat{\Pi})H,$$

which uses the unrestricted least squares estimates of the reduced form parameter Π_0.

Tests of the identification of γ can then be based on the smallest characteristic roots of the sample matrices $R_n(\tilde{\pi}_0)$ or $R_n^*(\hat{\pi})$, and the preceding two sections have established a limit normal distribution for these characteristic roots under fairly general conditions.

Before considering test statistics, it is necessary to discuss the nature of the hypotheses for which test statistics are to be provided. It seems sensible, from both theoretical and practical considerations, to propose that the null hypothesis be that "γ is identified", implying that $R_n(\pi_0)$ and $R_n^*(\pi_0)$ are nonsingular, and hence that their smallest characteristic roots are positive. The alternative hypothesis would then be that "γ is not identified", so that $R_n(\pi_0)$ and $R_n^*(\pi_0)$ are singular, with at least one zero root. Justification for this viewpoint will shortly be given, but it is interesting to note that the Farebrother and Savin [1974] "single root test" has "lack of identification" as null hypothesis, as does the Koopmans and Hood [1953] "double root test": this was noted in subsection 4.2.2.

Whether one accepts this choice of null hypothesis depends on one's attitude: theoreticians who hope to gain kudos by pointing out that certain estimated models seem to correspond to unidentified population structures might not agree with this choice. On the other hand, one would imagine that most practitioners facing the daunting task of building and maintaining "reasonable" models would presumably like to be protected against saying that their model is unidentified, when it is not, too often. That is, they would wish to treat "γ is identified" as the null hypothesis, given the lack of symmetry in the roles of the null and alternative hypotheses in classical hypothesis testing.

It is interesting that a parallel with this situation occurs even in the linear model: most applied workers use t-statistics (or even an F-statistic) with a view to <u>rejecting</u> the null hypothesis that a regression parameter is zero, on the grounds that their theory leading to the inclusion of the corresponding variable in the regression is confirmed. Yet, the very same practitioners would wish to confirm a theory of linear restrictions amongst the regression parameters by <u>accepting</u> the null hypothesis that the restrictions hold. Both of these views are seeking to confirm the specification of the assumed model.

It is also interesting that the nature of the test statistics considered determine to some extent the appropriate null hypothesis, although it will turn out to be possible to derive a test of sorts for either null hypothesis.

<u>4.6.2.</u> Given that the null hypothesis is taken to be the identification of γ, so that the smallest characteristic

roots of $R_n(\pi_0)$ and $R_n^*(\pi_0)$ are positive, it is convenient to make the simplification that the roots of $R_n(\pi_0)$ and $R_n^*(\pi_0)$ are all distinct: so the characteristic roots are

$\upsilon_{in}, \quad \upsilon_{in}^*, \quad i = 1,\ldots,q_0.$

This leads in turn to considerable simplification in the limit distribution structure of

$n^{1/2}(\tilde{\upsilon}_{in} - \upsilon_{in})$ and $n^{1/2}(\hat{\upsilon}_{in}^* - \upsilon_{in}^*), \quad i = 1,\ldots,q_0,$

where $\tilde{\upsilon}_{in}$, $\hat{\upsilon}_{in}^*$ are the roots of $R_n(\tilde{\pi}_0)$ and $R_n^*(\hat{\pi})$ respectively. Let the canonical decompositions of $R_n(\pi_0)$ and $R_n^*(\pi_0)$ be

$R_n(\pi_0) = \Gamma_n \Upsilon_n \Gamma_n', \quad R_n^*(\pi_0) = \Gamma_n^* \Upsilon_n^* \Gamma_n^{*\prime},$

with the corresponding limiting versions being

$R(\pi_0) = \Gamma\Upsilon\Gamma', \quad R^*(\pi_0) = \Gamma^*\Upsilon^*\Gamma^{*\prime};$

then, one can write, analogously to equations <4.5.5.2> and <4.5.5.3>,

$\Gamma_n = [g_{1n},\ldots,g_{q_0n}], \quad \Gamma_n^* = [g_{1n}^*,\ldots,g_{q_0n}^*],$

$\Gamma = [g_1,\ldots,g_{q_0}], \quad \Gamma^* = [g_1^*,\ldots,g_{q_0}^*].$

Correspondingly, the characteristic vectors of $R_n(\tilde{\pi}_0)$ and $R_n^*(\hat{\pi})$ are denoted

$\tilde{\mathcal{K}}_n = [\tilde{k}_{1n},\ldots,\tilde{k}_{q_0n}],$

$\hat{\mathcal{K}}_n^* = [\hat{k}_{1n}^*,\ldots,\hat{k}_{q_0n}^*].$

It will also be convenient for summary purposes to define matrices $\mathcal{N}$ and $\mathcal{N}^*$ such that

$\Psi(v[R_n(\tilde{\pi}_0)];\theta^0) = L_{q_0}S_{q_0}\mathcal{N}S_{q_0}L_{q_0}',$

$\Psi(v[R_n^*(\hat{\pi})];\theta^0) = L_{q_0}S_{q_0}\mathcal{N}^*S_{q_0}L_{q_0}',$

using equations <4.4.1.11> and <4.4.2.2>: since the exact nature of $\mathcal{N}$ and $\mathcal{N}^*$ play no role in what follows, it seems redundant to repeat the complex expressions they represent. Thus,

$$n^{\frac{1}{2}}v(R_n(\tilde{\pi}_o) - R_n(\pi^0)) \overset{a}{\approx} N(0, \mathbf{L}_{\mathbf{q}_o}\mathbf{S}_{\mathbf{q}_o}\mathcal{N}\mathbf{S}_{\mathbf{q}_o}\mathbf{L}'_{\mathbf{q}_o}),$$
$$n^{\frac{1}{2}}v(R_n^*(\hat{\pi}) - R_n^*(\pi^0)) \overset{a}{\approx} N(0, \mathbf{L}_{\mathbf{q}_o}\mathbf{S}_{\mathbf{q}_o}\mathcal{N}^*\mathbf{S}_{\mathbf{q}_o}\mathbf{L}'_{\mathbf{q}_o});$$

estimates of the matrices $\mathcal{N}$ and $\mathcal{N}^*$, obtained by using FIML estimates in a finite sample version of $\mathcal{N}$, and least squares estimates in a finite sample version of $\mathcal{N}^*$, will be denoted

$\tilde{\mathcal{N}}_n$, $\hat{\mathcal{N}}_n^*$

for notational simplicity.

Then, from subsection 4.5.3. and in particular equation <4.5.3.2>,

$$n^{\frac{1}{2}}(\tilde{v}_{in} - v_{in}) \overset{a}{\approx} (\mathbf{g}_i' \otimes \mathbf{g}_i')\mathbf{D}'_{\mathbf{q}_o}n^{\frac{1}{2}}v(R_n(\tilde{\pi}_o) - R_n(\pi^0))$$
$$\overset{a}{\approx} N(0, (\mathbf{g}_i' \otimes \mathbf{g}_i')\mathcal{N}(\mathbf{g}_i \otimes \mathbf{g}_i)), \qquad \text{<4.6.2.1>}$$
$$n^{\frac{1}{2}}(\hat{v}_{in}^* - v_{in}^*) \overset{a}{\approx} (\mathbf{g}_i^{*\prime} \otimes \mathbf{g}_i^{*\prime})\mathbf{D}'_{\mathbf{q}_o}n^{\frac{1}{2}}v(R_n^*(\hat{\pi}) - R_n^*(\pi^0))$$
$$\overset{a}{\approx} N(0, (\mathbf{g}_i^{*\prime} \otimes \mathbf{g}_i^{*\prime})\mathcal{N}^*(\mathbf{g}_i^* \otimes \mathbf{g}_i^*)),$$

making repeated use of the results <1.6.3.7> and <1.6.3.3>. From these two marginal distributions, it is easy to construct the joint distribution of any collection of the sampling errors.

Estimates of the limiting variances are given by

$$\tilde{h}_{in}^2 = (\tilde{\mathbf{k}}_{in}' \otimes \tilde{\mathbf{k}}_{in}')\tilde{\mathcal{N}}_n(\tilde{\mathbf{k}}_{in} \otimes \tilde{\mathbf{k}}_{in}), \qquad i = 1,\ldots,q_o, \qquad \text{<4.6.2.2>}$$
$$\hat{h}_{in}^{*2} = (\hat{\mathbf{k}}_{in}^{*\prime} \otimes \hat{\mathbf{k}}_{in}^{*\prime})\hat{\mathcal{N}}_n^*(\hat{\mathbf{k}}_{in}^* \otimes \hat{\mathbf{k}}_{in}^*), \qquad i = 1,\ldots,q_o,$$

using the characteristic vectors $\tilde{\mathbf{k}}_{in}$, $\hat{\mathbf{k}}_{in}^*$ of $R_n(\tilde{\pi}_o)$ and $R_n^*(\hat{\pi})$ respectively and the natural estimators $\tilde{\mathcal{N}}$ and $\hat{\mathcal{N}}^*$ of $\mathcal{N}$ and $\mathcal{N}^*$.

<u>4.6.3.</u> In the following discussion, hypotheses and test statistics will be constructed and discussed using only the criterion matrix $R_n(\tilde{\pi}_o)$, but it is straightforward to simply

change the notation and make the discussion directly applicable to the criterion matrix $R_n^*(\hat{\pi})$.

Given the discussion in subsection 4.6.1., the null hypothesis of a test of the rank of $R_n(\pi_0)$ (equivalently, a test of the identification of γ) is

$$\mathbf{H}_0: \quad \upsilon_{\mathbf{q}_0 n} > 0, \tag{4.6.3.1}$$

whilst the simplest possible alternative is that the rank deficiency of $R_n(\pi_0)$ is one:

$$\mathbf{H}_1: \quad \upsilon_{\mathbf{q}_0 n} = 0. \tag{4.6.3.2}$$

However, since asymptotic theory is used to provide a distribution for the test statistic under this null hypothesis, the same test statistic will serve to test

$$\mathbf{H}_0': \quad \upsilon_{\mathbf{q}_0} > 0$$

against

$$\mathbf{H}_1': \quad \upsilon_{\mathbf{q}_0} = 0:$$

note that a basic assumption of the analysis in this Chapter is that

$$\upsilon_{in} \longrightarrow \upsilon_i, \quad i = 1,\ldots,\mathbf{q}_0.$$

It can be seen from the hypotheses <4.6.3.1> and <4.6.3.2> that the null hypothesis is composite in a rather nasty way: it seems clear that a one-sided test is required, to cope with the essential non-negativity of $\upsilon_{\mathbf{q}_0 n}$. A device often mentioned in textbooks for this type of problem is to choose a typical value, $\upsilon_{\mathbf{q}_0 n}^* > 0$ say, and test the hypotheses

$$\mathbf{H}_0^*: \quad \upsilon_{\mathbf{q}_0 n} = \upsilon_{\mathbf{q}_0 n}^*$$

$$\mathbf{H}_1^*: \quad \upsilon_{\mathbf{q}_0 n} < \upsilon_{\mathbf{q}_0 n}^*.$$

However, given the nature of the null hypothesis <4.6.3.1>, rejection of H_0^* simply because $v_{\mathbf{q}_0 n}$ appears to be less than $v^*_{\mathbf{q}_0 n}$ may not be relevant unless $v^*_{\mathbf{q}_0 n}$ is some suitably small number, such as the computational "tolerance factor" on the diagonal elements of a square matrix to be inverted. This idea does however suggest another possibility: the set of values $v^*_{\mathbf{q}_0 n}$ for which the hypothesis

$$v_{\mathbf{q}_0 n} \geqq v^*_{\mathbf{q}_0 n}$$

is not rejected forms a confidence set with confidence coefficient $1 - \alpha$, if α is the size of the test (see Lehmann [1959, pp78-83]; Aitchison [1964]). One can turn the idea around by forming a lower confidence bound for $v_{\mathbf{q}_0 n}$ and declaring the null hypothesis <4.6.3.1> to be accepted if the confidence bound is positive. If the confidence bound is non-positive, the decision is not quite so clear-cut: a cautious procedure would be to reject the null hypothesis.

It is clear from this argument that the procedure being proposed belongs to the type of reasonably informal significance testing procedure suggested in a different context by Brown, Durbin and Evans [1975]. "Clearcut evidence" in favour of the null hypothesis is being demanded in order to accept this hypothesis; if anything less than this occurs, it is suggested that the null hypothesis <4.6.3.1> be rejected.

Constructing the lower confidence bound is straightforward: by <4.6.2.1> and <4.6.2.2>,

$\tilde{h}^{-1}_{q_0n}n^{1/2}(\tilde{v}_{q_0n} - v_{q_0n}) \xrightarrow{d} w \sim N(0,1)$;

let a_{q_0} be the desired size of test, and let $w(a_{q_0})$ be the value such that

$P[w \leq w(a_{q_0})] = 1 - a_{q_0}$.

Then,

$P\{\tilde{h}^{-1}_{q_0n}n^{1/2}(\tilde{v}_{q_0n} - v_{q_0n}) \leq w(a_{q_0})\} \to 1 - a_{q_0}$,

or,

$P\{\tilde{v}_{q_0n} - n^{-1/2}\tilde{h}_{q_0n}w(a_{q_0}) \leq v_{q_0n}\} \to 1 - a_{q_0}$,

and it follows that

$\tilde{v}_{q_0n} - n^{-1/2}\tilde{h}_{q_0n}w(a_{q_0})$

is the desired lower confidence bound, with asymptotic confidence coefficient $1 - a_{q_0}$.

4.6.4. A somewhat more general situation is where it is assumed that if the null hypothesis <4.6.3.1> fails, the minimum rank of $R_n(\pi_0)$ or $R^*_n(\pi_0)$ is $\mathbf{h}$, so that one should test the pair of hypotheses

$\mathbf{H}_0$: $v_{q_0n} > 0$ <4.6.3.1>

$\mathbf{H}_1$: $v_{hn} > 0;\ v_{h+1,n} = 0, \ldots, v_{q_0n} = 0.$ <4.6.4.1>

Following the argument of the preceding subsection, one possibility might be to use the sum of the corresponding sample characteristic roots,

$\sum_{i=h+1}^{q_0} \tilde{v}_{in}$,

and apply the one-sided confidence bound procedure, where the estimated variance expression is, from <4.6.2.1> and <4.6.2.2>,

$\sum_{i=h+1}^{q_0} \sum_{j=h+1}^{q_0} (\tilde{k}'_{in} \otimes \tilde{k}'_{in})\tilde{\mathcal{N}}_n(\tilde{k}_{jn} \otimes \tilde{k}_{jn})$.

Another possibility would be to compare the $\mathbf{q}_0 - \mathbf{h}$ confidence bounds implicit in

$$P\{\tilde{v}_{in} - n^{-1/2}\tilde{h}_{in}w(a_i) \leq v_{in}\} \to 1 - a_i, \qquad \text{<4.6.4.2>}$$

for $i = \mathbf{h}+1,\ldots,\mathbf{q}_0$, where, if $w \sim N(0,1)$,

$$P\{w \leq w(a_i)\} = 1 - a_i.$$

A Bonferroni bound can then be used for the overall confidence coefficient, say, $1 - \mathbf{a}$:

$$1 - \mathbf{a} \geq 1 - \sum_{i=\mathbf{h}+1}^{\mathbf{q}_0} a_i \qquad \text{<4.6.4.3>}$$

(see e.g., Seber [1977, pp125–127]); $\mathbf{a}$ may then be interpreted as the overall size of the test of the null hypothesis <4.6.3.1> induced by this procedure. The other aspect of this Bonferroni argument is that it allows, in a rather crude way, for the dependence between the successive tests or confidence bounds.

This procedure is somewhat conservative, and has the unpleasant feature that decreasing a_i to ensure that

$$1 - \mathbf{a} = 0.95,$$

say, will have the effect of increasing $w(a_i)$, thus possibly increasing the chance that the corresponding confidence bound is negative.

4.6.5. A possibly countervailing advantage, however, of the confidence bound procedure is the ease with which a sequential test of rank may be performed; this is also possible with the sum of roots statistic

$$\sum_{i=\mathbf{h}+1}^{\mathbf{q}_0} \tilde{v}_{in},$$

but more calculation is required at each stage.

The basic idea of a sequential test of rank is to say that if the minimum rank of $R_n(\pi_o)$ under the alternative hypothesis <4.6.4.1> is $\mathbf{h}$, and the overall null hypothesis <4.6.3.1> is rejected, what can be concluded about the rank of $R_n(\pi_o)$ (or $R_n^*(\pi_o)$) ? The method adopted for answering this question is the "nested test principle": see for example, Darroch and Silvey [1963], Seber [1966], Mizon [1977], and in a somewhat different context, Anderson [1971].

Define the sets

$$\mathcal{J}_i^* = \{\nu_{\mathbf{h}+1,n} > 0,\ldots,\nu_{\mathbf{h}+i,n} > 0, \nu_{\mathbf{h}+i+1,n} \geqq 0,\ldots,\nu_{\mathbf{q}_o n} \geqq 0\}$$

$$i = 1,\ldots,\mathbf{q}_o - \mathbf{h},$$

so that

$$\mathcal{J}_{i+1}^* \subseteq \mathcal{J}_i^*,$$

and the associated hypothesis

$$\mathbf{H}_{io}^*: \quad \nu_{\mathbf{h}+i,n} > 0, \nu_{\mathbf{h}+i+1,n} \geqq 0,\ldots,\nu_{\mathbf{q}_o n} \geqq 0, \quad i = 1,\ldots,\mathbf{q}_o-\mathbf{h} \qquad \langle 4.6.5.1\rangle$$

asserts that the rank of $R_n(\pi_o)$ is at least $\mathbf{h}+i$. A test of $\mathbf{H}_{io}^*$ against the alternative

$$\mathbf{H}_{i1}^*: \quad \nu_{\mathbf{h}+i-1,n} > 0, \ \nu_{\mathbf{h}+i,n} = 0, \quad i = 1,\ldots,\mathbf{q}_o-\mathbf{h}$$

is then a test that rank $R_n(\pi_o)$ is at least $\mathbf{h}+i$.

Since the sets $\mathcal{J}_i^*$ are nested, the standard nested test procedure can be applied (see the references above), testing $\mathbf{H}_{1o}^*$ against $\mathbf{H}_{11}^*$, and proceeding to test $\mathbf{H}_{2o}^*$ against $\mathbf{H}_{21}^*$ only if $\mathbf{H}_{1o}^*$ is accepted, and so on. If at any stage, $\mathbf{H}_{io}^*$ is rejected, the conclusion that rank $R_n(\pi_o)$ is $\mathbf{h}+i-1$ is drawn; only if all the hypotheses $\mathbf{H}_{io}^*$, $i = 1,\ldots,\mathbf{q}_o-\mathbf{h}$ are accepted can the overall null hypothesis <4.6.3.1> be accepted.

Using the separate confidence bounds of equation <4.6.4.2> for $\nu_{h+1,n},\ldots,\nu_{h+i,n}$ to test H^*_{io} above, equation <4.6.5.1>, is tantamount to testing only

H_{io}: $\nu_{h+i,n} > 0$

against

H_{i1}: $\nu_{h+i,n} = 0$,

given the truth of $H^*_{i-1,o}$, since the hypothesis that $\nu_{h+1,n} > 0,\ldots,\nu_{h+i-1,n} > 0$ must have been accepted before attempting to test H^*_{io}.

The sum of roots test, using

$\sum_{i=h+1}^{q_o} \tilde{\nu}_{in}$

does not seem very sensible in this context, for, if $H^*_{i-1,o}$ is true, then

$\sum_{j=h+1}^{q_o-1} \nu_{jn} > 0$

and hence so will be

$\sum_{j=h+1}^{q_o} \nu_{jn}$.

Another possible procedure is to test the null hypothesis <4.6.3.1>,

$H_o = H_{q_o o}$: $\nu_{q_o n} > 0$

against the overall alternative <4.6.4.1>,

$H_1 = H_{q_o 1}$: $\nu_{q_o n} = 0$,

and then proceed to test

$H_{q_o-1,o}$: $\nu_{q_o-1,n} > 0$

against

$H_{q_0-1,1}$: $\nu_{q_0-1,n} = 0$

only if $H_0 = H_{q_0 0}$ has been rejected. This is the procedure described by Mizon [1977] as the "reverse" procedure, in contrast to the nested test procedure, which may be called the "forward" procedure. In this latter case, the separate confidence intervals for ν_{jn} are examined in the order $j = h+1,\ldots q_0$, whilst in the reverse procedure, the order would be $j = q_0, q_0-1,\ldots,h+1$. Given the remarks in subsection 4.6.3. concerning the interpretation of a non-positive confidence bound, this reverse procedure may be regarded as somewhat less straightforward than the forward procedure. In any case, even in more standard inferential situations, there is no reason why the same inferences should be obtained from each procedure.

The confidence bounds associated with the statistics based on the sum of the appropriate characteristic roots can be used in a "reverse" procedure for a sequential test of rank, for $H_{q_0-j,0}$ would be examined only on the assumption that

$H_{q_0-j+1,0},\ldots,H_{q_0 0} = H_0$

fail. Then,

$\sum_{j=q_0-i}^{q_0} \nu_{jn} = \nu_{q_0-i}$,

given the truth of $H_{q_0-i-1,0}$; thus, one would examine the sequence of confidence bounds associated with

$\tilde{\nu}_{q_0 n}$; $\tilde{\nu}_{q_0 n}+\tilde{\nu}_{q_0-1,n}$;...; $\tilde{\nu}_{q_0 n}+ \ldots +\tilde{\nu}_{h+1,n}$,

and declare the rank of $R_n(\pi_0)$ to be q_0-i if the hypothesis $H_{q_0-i,0}$ is accepted, but $H_{q_0-i-1,0}$ is rejected.

The Bonferroni bound on the overall confidence coefficient 1 - a, equation <4.6.4.3>, still applies, no matter which of the sequential procedures is used.

4.6.6 It is worth emphasising here that the tests proposed are essentially "system" tests of identification, and hence, if the overall null hypothesis of identification of γ fails, it will be impossible to declare that the source of the failure lies in a particular equation or group of equations. Such information might, however, be very helpful to an investigator seeking to eradicate a failure of identification from an hypothesised model. However, in the "usual special case" of within equation exclusion restrictions, a limiting distribution based on equation <4.5.5.6> is possible, given the assumption of distinct non-zero roots of the population matrix $R^*_n(\pi_0)$, and hence for its diagonal blocks $R^*_{in}(\pi_0)$, $i = 1,\ldots,m$. One can then use the procedures outlined in the preceding subsections to test the hypothesis that the smallest root of each $R^*_{in}(\pi_0)$ in turn is positive. It is worth pointing out that the null hypothesis of the identification of each equation of the model is the appropriate one, given the way in which the population characteristic roots appear in the limiting covariance matrix of equation <4.5.5.6>.

4.6.7. In subsection 4.6.2. it was assumed that the non-zero characteristic roots of $R_n(\pi_0)$ and $R^*_n(\pi_0)$ were distinct: in this subsection, this assumption is relaxed, and the implications for tests of rank discussed. In this case, there are t distinct population roots

$$\upsilon_{1n} > \dots > \upsilon_{tn},$$

with respective multiplicities $h_1,\dots,h_t$: then, tests of the rank of $R_n(\pi_0)$ can only proceed in steps of h_i sample characteristic roots at a time. Thus, if the null and alternative hypotheses are

$$H_0: \quad \upsilon_{tn} > 0$$

$$H_1: \quad \upsilon_{tn} = 0,$$

the latter hypothesis corresponds to the statement that

$$\text{rank } R_n(\pi_0) = \sum_{i=1}^{t-1} h_i = p_{t-1}.$$

(see equation <4.5.1.3>).

The h_t smallest sample roots

$$\tilde{\upsilon}_{p(t-1)+1,n},\dots,\tilde{\upsilon}_{q_0 n}$$

correspond to the smallest population root υ_{tn}, so it would be natural to use as the basis of a confidence region test the average of these sample roots (compare Anderson [1963]):

$$h_t^{-1} \sum_{j=p+1}^{q_0} \tilde{\upsilon}_{jn}.$$

An estimate of the variance of this expression is very tedious and long-winded to write out, being based very closely on the combination of equation <4.5.3.2> with equations <4.4.1.11> and <4.4.2.2>: it is proposed simply to denote this estimate by

$$h_t^{-2} \tilde{\Psi}_{tn}$$

for the criterion $R_n(\tilde{\pi}_0)$ and by

$$h_t^{-2} \tilde{\Psi}^*_{tn}$$

for $R_n^*(\hat{\pi})$. This notation is a little non-standard, but the reward here for sticking to the notational principles is zero.

Thus, following through the lower confidence bound argument of subsection 4.6.3. yields

$$\sum_{j=p-1}^{q_0} \tilde{v}_{jn} - n^{-\frac{1}{2}} h_t \tilde{\Psi}_{tn}^{-\frac{1}{2}} \omega(a) \leq v_{tn},$$

which will have asymptotic confidence coefficient $1 - a$.

It seems unlikely that an investigator will have sufficient prior knowledge of the multiplicities of the population roots v_{in}, v^*_{in} required for this analysis, so that the procedure is unlikely to be very useful; it is included only for completeness.

4.7. Tests of Consistency

4.7.1. In subsection 4.1.2., it was observed that there are two conditions that are jointly necessary and sufficient for the identification of the parameter vector γ in the simultaneous equations model defined by equations <3.1.3.5> and <3.1.3.6>:

$$(I_m \otimes Z_1)g_0 = u_0,$$

$$g_0 = H\gamma + h,$$

namely, that there exists a solution to

$$(I_m \otimes \Omega_0)(H\gamma + h) = 0$$

and that the matrix

$$(I_m \otimes \Omega_0)H$$

has full column rank q_0. These conditions have been described as the "consistency condition" and the "rank condition". In section 4.6., tests of the rank condition were described, under the assumption that the consistency condition held; in this section, tests of the consistency condition are proposed, of a similar kind to the tests of the rank condition.

It is clear that satisfaction of the consistency condition is logically prior to the satisfaction of the rank condition, for if no solution to

$$(I_m \otimes \Omega_0)(H\gamma + h) = 0$$

exists, whether or not

$$\text{rank } (I_m \otimes \Omega_0)H = q_0$$

is irrelevant.

The consistency condition can be expressed as

$\text{rank}[(I_m \otimes Q_0)h : (I_m \otimes Q_0)H] = \text{rank } (I_m \otimes Q_0)H$:

let

$H^* = [h : H]$;

then, the condition is equivalent to saying that

$(I_m \otimes Q_0)H^*$

has exactly one linearly dependent column, provided that

$\text{rank } (I_m \otimes Q_0)H = q_0$.

Proceeding by analogy with the tests of rank proposed in the preceding section, a natural way to test the consistency condition would be to examine the smallest characteristic roots of matrices like

$$\mathcal{S}_n(\pi_0) = n^{-1}H^{*\prime}(\Sigma_0^{-1} \otimes Q_0'X'XQ_0)H^* \qquad \langle 4.7.1.1\rangle$$

or

$$\mathcal{S}_n^*(\pi_0) = n^{-1}H^{*\prime}(I_m \otimes Q_0'X'XQ_0)H^*, \qquad \langle 4.7.1.2\rangle$$

which are clearly closely related to the matrices $R_n(\pi_0)$ and $R_n^*(\pi_0)$ of equations <4.4.1.1> and <4.4.1.2>. The corresponding sample matrices are

$\mathcal{S}_n(\tilde{\pi}_0) = n^{-1}H^{*\prime}(\tilde{\Sigma}_0^{-1} \otimes \tilde{Q}_0'X'X\tilde{Q}_0)H^*$,

using the FIML estimators of γ, Π_0 and Σ_0, and

$\mathcal{S}_n^*(\hat{\pi}) = n^{-1}H^{*\prime}(I_m \otimes \hat{Q}'X'X\hat{Q})H^*$,

which simply uses the least squares estimator of Π_0,

$\hat{\Pi} = (X'X)^{-1}X'Y$.

Note that $\mathcal{S}_n(\pi_0)$ and $\mathcal{S}_n^*(\pi_0)$ have dimensions $q_0 + 1$ by $q_0 + 1$.

The population and sample characteristic roots will be denoted

v_{in}, $\tilde{v}_{in}$, v^*_{in}, and $\hat{v}^*_{in}$

for

$\mathcal{S}_n(\pi_0)$, $\mathcal{S}_n(\tilde{\pi}_0)$, $\mathcal{S}^*_n(\pi_0)$ and $\mathcal{S}^*_n(\hat{\pi})$

respectively. The limit normal distributions for $R_n(\tilde{\pi}_0)$ and $R^*_n(\hat{\pi})$ obtained in subsection 4.4.1. (equations <4.4.1.10> and <4.4.1.11>) and 4.4.2. (equations <4.4.2.1> and <4.4.2.2>) can be carried over directly, by appropriately replacing H by H^* in these expressions. More precisely, H^* has to be partitioned as H was in equation <4.4.1.6>:

$$H^{*\prime} = [H_1^{*\prime},\ldots,H_m^{*\prime}], \qquad \text{<4.7.1.3>}$$

and then as in equation <4.4.1.8>,

$$H_i^{*\prime} = [H_{i.1}^{*\prime} : H_{i.2}^{*\prime}]; \qquad \text{<4.7.1.4>}$$

the dimensionality of the matrices $\mathbf{S}_{\mathbf{q}_0}$ and $\mathbf{L}_{\mathbf{q}_0}$ has to be changed as well, so that they become $\mathbf{S}_{\mathbf{q}_0+1}$, $\mathbf{L}_{\mathbf{q}_0+1}$.

The matrices of characteristic vectors of $\mathcal{S}_n(\pi_0)$ and $\mathcal{S}^*_n(\pi_0)$ will be denoted $\mathcal{C}_n$ and $\mathcal{C}^*_n$, with columns $\mathbf{c}_{in}$, $\mathbf{c}^*_{in}$:

$$\mathcal{C}_n = [\mathbf{c}_{1n},\ldots,\mathbf{c}_{\mathbf{q}_0+1,n}],$$

$$\mathcal{C}^*_n = [\mathbf{c}^*_{1n},\ldots,\mathbf{c}^*_{\mathbf{q}_0+1,n}],$$

whilst the characteristic vectors of the sample matrices $\mathcal{S}_n(\tilde{\pi})$ and $\mathcal{S}^*_n(\hat{\pi})$ will be denoted $\tilde{\mathcal{C}}_n$, $\hat{\mathcal{C}}^*_n$, respectively, with columns $\tilde{\mathbf{c}}_{in}$, $\hat{\mathbf{c}}^*_{in}$.

The covariance matrices of the limit normal distributions of

$$n^{1/2}v(\mathcal{S}_n(\tilde{\pi}_0) - \mathcal{S}_n(\pi^0))$$

and

$$n^{1/2}v(\mathcal{S}^*_n(\hat{\pi}) - \mathcal{S}^*_n(\pi^0))$$

can then be written as

$$\Psi[v(S_n(\tilde{\pi}_0));\theta^0] = L_{q_0+1}S_{q_0+1}\mathcal{B}S_{q_0+1}L'_{q_0+1} \qquad \langle 4.7.1.5\rangle$$

and

$$\Psi[v(S_n^*(\hat{\pi}));\theta^0] = L_{q_0+1}S_{q_0+1}\mathcal{B}^*S_{q_0+1}L'_{q_0+1} \qquad \langle 4.7.1.6\rangle$$

respectively, where

$$\begin{aligned}\mathcal{B} = 4\sum_{i=1}^{m}\sum_{j=1}^{m}&[H^{*\prime}_{i.1}A^{0-1\prime} \otimes H^{*\prime}(\Sigma^{0-1}e_i \otimes \Pi^{0\prime}M_X\Pi^0)]\\ &\times H[H'(\Sigma^{0-1} \otimes \Pi^{0\prime}M_X\Pi^0)H]^{-1}H'\\ &\times [A^{0-1}H^*_{j.1} \otimes (e_j'\Sigma^{0-1} \otimes \Pi^{0\prime}M_X\Pi^0)H^*]; \qquad \langle 4.7.1.7\rangle\end{aligned}$$

$$\mathcal{B}^* = 4\sum_{i=1}^{m}\sum_{j=1}^{m}[H^{*\prime}_{i.1}\Omega^0H^*_{j.1} \otimes H^{*\prime}(e_ie_j' \otimes \Pi^{0\prime}M_X\Pi^0)H^*]. \qquad \langle 4.7.1.8\rangle$$

4.7.2. The null hypothesis that the consistency condition holds is equivalent to

H_0: $v_{q_0+1,n} = 0$ (or $v^*_{q_0+1,n} = 0$)

with the alternative

H_1: $v_{q_0+1,n} > 0$ (or $v^*_{q_0+1,n} > 0$).

In short, under the null hypothesis, $S_n(\pi_0)$ or $S_n^*(\pi_0)$ should have a single zero population root. It will then follow from equation <4.6.2.1> that

$n^{1/2}(\tilde{v}_{q_0+1,n} - v_{q_0+1,n}) \overset{a}{\sim} N(0,(c'_{q_0+1} \otimes c'_{q_0+1})\mathcal{B}(c_{q_0+1} \otimes c_{q_0+1}))$,

with an analogous result for $\hat{v}^*_{q_0+1}$. Replacing $\mathcal{B}$ by its estimate using the FIML estimates, and c_{q_0+1} by $\tilde{c}_{q_0+1,n}$ will enable a perfectly standard one-sided "asymptotic" t-test to be performed.

4.7.3. One can construct a variant of this idea for the special case where the matrix H is block diagonal (corresponding to "within equation" restrictions only), and for which the null hypothesis of consistency will require the smallest roots of $S_n(\pi_0)$ or $S_n^*(\pi_0)$ to be zero. For, when

$$H = \mathop{\oplus}_{i=1}^{m} H_{ii},$$

the equation system

$$(I_m \otimes \Omega_0)H\gamma = -(I_m \otimes \Omega_0)h$$

is clearly equivalent to

$$\Omega_0 H_{ii}\gamma_{.i} = -\Omega_0 h_{.i}, \qquad i = 1,\ldots,m.$$

The consistency of this collection of equations can be checked via the number of linearly independent columns of the block diagonal matrix

$$(I_m \otimes \Omega_0)H^*,$$

where

$$H^* = \mathop{\oplus}_{i=1}^{m} H_{ii}^*,$$

$$H_{ii}^* = [h_{.i} : H_{ii}].$$

This leads naturally, by the arguments of the preceding subsections, to criterion matrices of the form

$$T_n(\pi_0) = n^{-1}H^{*\prime}(\Sigma_0^{-1} \otimes \Omega_0'X'X\Omega_0)H^*, \tag{4.7.3.1}$$

$$T_n^*(\pi_0) = n^{-1}H^{*\prime}(I_m \otimes \Omega_0'X'X\Omega_0)H^*, \tag{4.7.3.2}$$

with sample versions

$$T_n(\tilde{\pi}_0) = n^{-1}H^{*\prime}(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Omega}_0'X'X\tilde{\Omega}_0)H^*,$$

$$T_n^*(\hat{\pi}) = n^{-1}H^{*\prime}(I_m \otimes \hat{\Omega}'X'X\hat{\Omega})H^*.$$

Denote the characteristic roots of $T_n(\pi_0)$ and $T_n^*(\pi_0)$ by f_{in}, f_{in}^*, $i = 1,\ldots,t$, where

$$t = q_0 + m,$$

if it is assumed that all the non-zero roots are distinct. The limit distribution of the distinct elements of the criterion matrices follows that given in subsection 4.7.1.

above, equations <4.7.1.5>–<4.7.1.8>, when $H^{\#}$ is partitioned to match the partitioning of H^{*} given by equations <4.7.1.3> and <4.7.1.4>, and $\mathbf{S}_{\mathbf{q}_0+1}$, $\mathbf{L}_{\mathbf{q}_0+1}$ are replaced by $\mathbf{S}_{\mathbf{q}_0+m}$, $\mathbf{L}_{\mathbf{q}_0+m}$.

The null hypothesis that the consistency condition holds then demands that the m smallest roots of $\mathcal{T}_n(\pi_0)$ and $\mathcal{T}_n^{*}(\pi_0)$ are zero: it is more convenient to write

$\mathbf{H}_0$: $\quad \mathbf{f}_{\mathbf{q}_0+1,n} = 0,\ldots,\mathbf{f}_{\mathbf{q}_0+m,n} = 0,$

$\mathbf{H}_1$: $\quad$ at least one of these is positive,

as the hypotheses to be tested, rather than defining the root $\mathbf{f}_{\mathbf{t}-1,n}$ to be equal to zero and have multiplicity m.

Using the limit distribution

$$m^{-1}\sum_{i=1}^{m}\tilde{\mathbf{f}}_{\mathbf{q}_0+i,n} \overset{a}{\sim} N(0,\mathbf{b}^2)$$

under the null hypothesis, a straightforward large sample t-test can be obtained. The reward for specifying carefully the long expression for this variance is small, so it seems better not to give it precisely, as in a similar situation in subsection 4.6.7. .

It is worth noting that even if the criterion $\mathcal{T}_n^{*}(\hat{\pi})$ is used, the consistency condition test is inevitably a system-wide test, because of the way in which the smallest characteristic roots of each diagonal block of $\mathcal{T}_n^{*}(\hat{\pi})$ are brought together in the test statistic. Thus, a "one equation at a time" test is impossible, despite the nature of the structural equations being assumed.

4.7.4 It is possible to construct analogous tests for models where the actual number of normalisation rules is somewhere between 1 and m, but the test statistics are quite complex, and little further insight can be obtained.

A joint test of identification can be obtained by testing for consistency and rank separately. For, the rank test is a test of identification only if consistency is assumed to hold, and equally, it is no use knowing that say $S_n(\pi_0)$ has a single zero root unless it corresponds to the "correct" linear combination of its columns - that is, unless the rank conditions holds. Such a joint test is induced by the pair of dependent tests, and thus, one would have to use the Bonferroni inequality to control the size of the joint test.

4.8. Some Comparisons and Interpretations

4.8.1. The existing tests discussed in section 4.2. were based on the use of the LIML estimator in a single overidentified structural equation, the rest of the model being assumed to be in reduced form. To compare the tests of sections 4.6. and 4.7. with these existing tests, it will be necessary to specialise the results of these two sections to the special situation in which the LIML estimators are used.

The "null hypothesis" simultaneous equations model defined by equations <3.1.3.5> and <3.1.3.6>,

$(I_m \otimes Z_1)g_0 = u_0,$

$g_0 = H\gamma + h,$

but this latter expression is equivalent, in the current situation, to

$$g_0 = \begin{bmatrix} c_{0.1} \\ g_{0.2} \end{bmatrix} = \begin{bmatrix} H_{11} & : & 0 \\ 0 & : & I_{m-1} \otimes \begin{bmatrix} 0_{mk_1} \\ -I_{k_1} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \gamma_{.1} \\ \gamma_{.2} \end{bmatrix} + \begin{bmatrix} h_{.1} \\ h_{.2} \end{bmatrix},$$

from equations <3.7.2.3>-<3.7.2.5>. The corresponding definition of H^* is

$$H^* = [h : H]$$

$$= \begin{bmatrix} h_{.1} & : & H_{11} & : & 0 \\ h_{.2} & : & 0 & : & I_{m-1} \otimes \begin{bmatrix} 0_{mk_1} \\ -I_{k_1} \end{bmatrix} \end{bmatrix}.$$

It is also worth noting that

$$(I_{m-1} \otimes \Omega_0)(I_{m-1} \otimes \begin{bmatrix} O_{mk_1} \\ -I_{k_1} \end{bmatrix}) = -(I_{m-1} \otimes I_{k_1})$$

In these circumstances, the criterion of equation <4.7.1.1> is

$$\begin{aligned} S_n(\pi_0) &= n^{-1}H^{*\prime}(\Sigma_0^{-1} \otimes \Omega_0'X'X\Omega_0)H^* \\ &= n^{-1}\begin{bmatrix} h_{.1}'\Omega_0' & : & h_{.2}'(I_{m-1} \otimes \Omega_0') \\ H_{11}'\Omega_0' & : & 0 \\ 0 & : & -(I_{m-1} \otimes I_{k_1}) \end{bmatrix}(\Sigma_0^{-1} \otimes X'X) \\ &\quad \times \begin{bmatrix} \Omega_0 h_{.1} & : \Omega_0 H_{11} : & 0 \\ (I_{m-1} \otimes \Omega_0)h_{.2} & : \quad 0 \quad : & -(I_{m-1} \otimes I_{k_1}) \end{bmatrix}, \end{aligned}$$

and one can deduce from this the form of the rank criterion matrix of equation <4.4.1.1>:

$$R_n(\pi_0) = n^{-1}\begin{bmatrix} H_{11}'\Omega_0' : & 0 \\ 0 & : -I_{m-1} \otimes I_{k_1} \end{bmatrix}(\Sigma_0^{-1} \otimes X'X)\begin{bmatrix} \Omega_0 H_{11} : & 0 \\ 0 & : -I_{m-1} \otimes I_{k_1} \end{bmatrix}$$

It is clear from this latter expression that questions about the rank of $R_n(\pi_0)$ reduce to questions about

$$\text{rank } n^{-1}[H_{11}'\Omega_0' : \ 0](\Sigma_0^{-1} \otimes X'X)\begin{bmatrix} \Omega_0 H_{11} \\ 0 \end{bmatrix} = \text{rank } n^{-1}H_{11}'\Omega_0'X'X\Omega_0 H_{11},$$

that is, the rank criterion of the overidentified equation. However, as noted in subsection 4.4.3., there is no simplification in the limit distribution of the sample criterion matrix $R_n(\tilde{\pi}_0)$ and hence that of the sample characteristic roots.

For the criterion matrix of equation <4.7.1.2>,

$$S_n^*(\pi_0) = n^{-1}H^{*\prime}(I_m \otimes \Omega_0'X'X\Omega_0)H^*$$

$$= n^{-1}\begin{bmatrix} h'(I_m \otimes \Omega_0'X'X\Omega_0)h : h_{.1}'\Omega_0'X'X\Omega_0H_{11} : -h_{.2}'(I_{m-1} \otimes \Omega_0'X'X) \\ H_{11}'\Omega_0'X'X\Omega_0h_{.1} \quad : H_{11}'\Omega_0'X'X\Omega_0H_{11} : \quad 0 \\ -(I_{m-1} \otimes X'X\Omega_0)h_{.2} : \quad 0 \quad : \quad I_{m-1} \otimes X'X \end{bmatrix},$$

from which the criterion matrix of equation <4.4.1.2> may be deduced:

$$R_n^*(\pi_0) = n^{-1}\begin{bmatrix} H_{11}'\Omega_0'X'X\Omega_0H_{11} : & 0 \\ 0 & : I_{m-1} \otimes X'X \end{bmatrix}.$$

Again, one can see that the rank of $R_n^*(\pi_0)$ is determined by the same matrix as in the case of the criterion $R_n(\pi_0)$, that is, by

$$R_{1n}^*(\pi_0) = n^{-1}H_{11}'\Omega_0'X'X\Omega_0H_{11}.$$

The limit distribution of the smallest root of $R_{1n}^*(\hat{\pi})$ was obtained in subsection 4.5.5. (given that the corresponding population root has multiplicity 1) and is given in equation <4.5.5.6>.

Since

$$\hat{\Omega}'X'X\hat{\Omega} = Z_1'P_XZ_1,$$

(where $P_X = X(X'X)^{-1}X'$), one can see that rank tests based on $R_n(\tilde{\pi})$ or $R_n^*(\hat{\pi})$ are similar in intent to the Farebrother and Savin [1974] "single root test" discussed in subsection 4.2.2., with hypotheses <4.2.2.6> and <4.2.2.7>. The only substantive difference is that for the tests proposed in this thesis, the roles of the null and the alternative hypotheses are reversed, in comparison with the null and alternative hypotheses adopted by Farebrother and Savin. Needless to say, the distributional properties of the two types of test statistic are different.

4.8.2 The consistency condition for the $m - 1$ equations which are expressed in reduced form can be shown to be automatically satisfied, by using the matrices $\mathcal{T}_n(\pi_0)$, $\mathcal{T}_n^*(\pi_0)$ defined in equations <4.7.3.1> and <4.7.3.2>, which use the block diagonal matrix

$$H^* = \underset{i=1}{\overset{m}{\oplus}} H^*_{ii}.$$

In the current circumstances,

$$H^*_{ii} = \left[h_{.2i} \;:\; \begin{bmatrix} 0 \\ -I_{k_1} \end{bmatrix}\right],$$

where

$$h'_{.2} = (h'_{.21}, \ldots, h'_{.2m}),$$

and

$$\Omega_0 H^*_{11} = [\Omega_0 h_{.2i} \;:\; -I_{k_1}], \qquad i = 2, \ldots, m.$$

One can therefore conclude that the consistency tests are examining the consistency condition solely for the parameters of the first, overidentified, equation; however, as with some of the rank tests, they are model wide in execution, in contrast to the "single root" tests of LIML estimation, based on the smallest characteristic root $l^*_{q_{01}+1}$ associated with equation <4.2.2.1>.

4.8.3. In subsection 4.2.2., it was argued that so long as the "rank condition"

$$l_{q_{01}} > 0$$

(see equation <4.2.2.6>) holds, the "single root" test is a test of the overidentifying restrictions on the first equation: it is worth considering whether a similar interpretation is possible for the consistency tests based on the smallest roots of $\mathcal{S}_n(\tilde{\pi}_0)$, $\mathcal{T}_n(\tilde{\pi}_0)$, $\mathcal{S}_n^*(\hat{\pi})$ or $\mathcal{T}_n^*(\hat{\pi})$.

Essentially, the consistency tests ask whether, given that the rank condition is met, there is a solution

$\gamma^{*\prime} = (1, \gamma^{\prime})$

to the limiting equations

$\mathcal{S}(\pi^0)\gamma^* = H^{*\prime}(\Sigma^{0-1} \otimes \Omega^{0\prime}M_x\Omega^0)H^*\gamma^* = 0$

(for example); as observed earlier, this is a meaningful question only when the model is overidentified. Thus, the question amounts to asking whether there exists a vector g_0 satisfying the restrictions

$g_0 = H\gamma + h$,

that is, whether the restrictions are true. Hence, the consistency tests may be similarly interpreted as tests of overidentifying restrictions; it will become clear that the test statistics are (generally) very different from the more classical tests of overidentifying restrictions which will be discussed in the next Chapter.

4.8.4. In concluding this Chapter, one might properly raise the question of whether the various tests proposed here are really feasible or operational. After all, most of them are system-wide tests, and thus require the calculation of the smallest characteristic roots of comparatively large matrices. In addition, the nature of the limiting distribution used to provide an approximate distribution for the sample characteristic roots used in the tests is rather complex and not particularly easy to handle.

Several comments may be offered in defence of the proposed tests. First, many FIML estimation programs have a singularity check on the "information matrix" of the problem, and the nature of this check varies from program to program; in some cases, the program will terminate if the check is not satisfactory. The rank tests allow the possibility of declaring that a model is unidentified even if the check is passed, and conversely, that the population model is identified even if the check fails. In short, some statistical inference on rank is possible. Secondly, the author's intention was to see what could be done to provide tests of identification, and what were the sorts of problems that might be encountered: such a viewpoint is not likely to be seen as particularly sensible by a practitioner, but it is none the less a valid "academic" viewpoint. Finally, although not heavily emphasised here, the rank and consistency tests are particularly straightforward for the "usual special case" type of simultaneous equations model estimated say by two-stage least squares: one could in fact integrate estimation and testing by using an algorithm based on a singular value decomposition of the relevant moment matrices.

Appendix to Chapter 4

4.A.1 This appendix derives equations <4.5.1.1.0>-<4.5.1.15> and defines certain matrices used in these equations, starting from equations <4.5.1.6> and <4.5.1.8>,

$$\mathcal{U}_n = n^{1/2}[\mathcal{E}_n(\Upsilon_n + n^{-1/2}\mathcal{W}_n)\mathcal{E}_n' - \Upsilon_n], \qquad \text{<4.A.1.1>}$$

and

$$\mathcal{E}_n\mathcal{E}_n' = I_r = {}_{i\,\underline{=}\,1}^{\;t}\!\!\oplus I_{h_i}. \qquad \text{<4.A.1.2>}$$

Partitioning up the matrices $\mathcal{E}_n$, $\mathcal{U}_n$ as

$\mathcal{E}_n = \|\mathcal{E}_{ijn}\|$, $\quad \mathcal{U}_n = \|\mathcal{U}_{ijn}\|$,

each block being $h_i \times h_j$,

$\mathcal{W}_n = {}_{i\,\underline{=}\,1}^{\;t}\!\!\oplus \mathcal{W}_{in}$,

the diagonal blocks being $h_i \times h_i$, with (equation <4.5.1.9>)

$\mathcal{W}_{in} = n^{1/2}(\hat{\Upsilon}_{in} - \upsilon_{in}I_{h_i})$.

Similarly,

$\Upsilon_n = {}_{i\,\underline{=}\,1}^{\;t}\!\!\oplus \Upsilon_{in}$,

with $\Upsilon_{in} = \upsilon_{in}I_{h_i}$.

It will also be helpful to define

$\mathcal{F}_{ijn} = n^{1/2}\mathcal{E}_{ijn}$, $\; i \neq j$.

Then, from equation <4.A.1.2>,

$0 = \sum_{k=1}^{t}\mathcal{E}_{ikn}\mathcal{E}_{jkn}'$, $\quad i \neq j$,

and, multiplying through by $n^{1/2}$, one obtains

$0 = \mathcal{E}_{iin}\mathcal{F}_{jin}' + \mathcal{F}_{ijn}\mathcal{E}_{jjn}' + n^{-1/2}\sum^{*}\mathcal{F}_{ikn}\mathcal{F}_{jkn}'$, $\quad i \neq j$,

where $\sum^{*}$ denotes summation over $k = 1,\dots,t$, but with $i \neq k$, $j \neq k$. Then, equation <4.5.1.15> is produced:

$$0 = \mathcal{E}_{iin}\mathcal{F}_{jin}' + \mathcal{F}_{ijn}\mathcal{E}_{jjn}' + n^{-1/2}\mathcal{L}_{ijn}, \quad i \neq j \qquad \text{<4.A.1.3>}$$

where

$$\mathcal{L}_{ijn} = \Sigma^{*}\mathcal{F}_{ikn}\mathcal{F}'_{jkn}. \qquad \text{<4.A.1.4>}$$

In addition,

$$I_{h_i} = \mathcal{E}_{iin}\mathcal{E}'_{iin} + n^{-1}\Sigma^{\ddagger}\mathcal{F}_{ikn}\mathcal{F}'_{ikn},$$

where the summation is over $k = 1,\ldots,t$, $k \neq i$; define

$$\mathcal{L}_{iin} = \Sigma^{\ddagger}\mathcal{F}_{ikn}\mathcal{F}'_{ikn}. \qquad \text{<4.A.1.5>}$$

From equation <4.A.1.2>,

$$\begin{aligned}\mathcal{U}_{iin} &= n^{1/2}[v_{in}\mathcal{E}_{iin}\mathcal{E}'_{iin} + n^{-1/2}\mathcal{E}_{iin}\mathcal{W}_{in}\mathcal{E}'_{iin} - v_{in}I_{h_i} \\ &\qquad + n^{-1}\Sigma^{\ddagger}\mathcal{F}_{ikn}(v_{kn}I_{h_k} + n^{-1/2}\mathcal{W}_{kn})\mathcal{F}'_{ikn}] \\ &= n^{1/2}[v_{in}I_{h_i} - n^{-1}v_{in}\Sigma^{\ddagger}\mathcal{F}_{ikn}\mathcal{F}'_{ikn} + n^{-1/2}\mathcal{E}_{iin}\mathcal{W}_{in}\mathcal{E}'_{iin} - v_{in}I_{h_i} \\ &\qquad + n^{-1}\Sigma^{\ddagger}\mathcal{F}_{ikn}(v_{kn}I_{h_k} + n^{-1/2}\mathcal{W}_{kn})\mathcal{F}'_{ikn}].\end{aligned}$$

Let

$$\mathcal{M}_{iin} = \Sigma^{\ddagger}v_{kn}\mathcal{F}_{ikn}\mathcal{F}'_{ikn}, \qquad \text{<4.A.1.6>}$$

$$\mathcal{Y}_{iin} = \Sigma^{\ddagger}\mathcal{F}_{ikn}\mathcal{W}_{kn}\mathcal{F}'_{ikn}; \qquad \text{<4.A.1.7>}$$

then, from equation <4.A.1.3>,

$$\mathcal{U}_{iin} = \mathcal{E}_{iin}\mathcal{W}_{in}\mathcal{E}'_{iin} + n^{-1/2}(\mathcal{M}_{iin} - v_{in}\mathcal{L}_{iin}) + n^{-1}\mathcal{Y}_{iin},$$

which is equation <4.5.1.10>; recall that for $i = t$, $v_{tn} \equiv 0$, so that equation <4.5.1.11> is produced.

Next,

$$\begin{aligned}\mathcal{U}_{ijn} &= n^{1/2}\sum_{k=1}^{t}\mathcal{E}_{ikn}(v_{kn}I_{h_k} + n^{-1/2}\mathcal{W}_{kn})\mathcal{E}'_{jkn} \\ &= n^{1/2}[v_{in}\mathcal{E}_{iin}\mathcal{E}'_{jin} + v_{jn}\mathcal{E}_{ijn}\mathcal{E}'_{jjn} + n^{-1/2}\mathcal{E}_{iin}\mathcal{W}_{in}\mathcal{E}'_{jin} \\ &\qquad + n^{-1/2}\mathcal{E}_{ijn}\mathcal{W}_{jn}\mathcal{E}'_{jjn} + \Sigma^{*}\mathcal{E}_{ikn}(v_{kn}I_{h_k} + n^{-1/2}\mathcal{W}_{kn})\mathcal{E}'_{jkn}] \\ &= v_{in}\mathcal{E}_{iin}\mathcal{F}'_{jin} + v_{jn}\mathcal{F}_{ijn}\mathcal{E}'_{jjn} + n^{-1/2}(\mathcal{E}_{iin}\mathcal{W}_{in}\mathcal{F}'_{jin} \\ &\qquad + \mathcal{F}_{ijn}\mathcal{W}_{jn}\mathcal{E}'_{jjn} + \Sigma^{*}v_{kn}\mathcal{F}_{ikn}\mathcal{F}'_{jkn}) + n^{-1}\Sigma^{*}\mathcal{F}_{ikn}\mathcal{W}_{kn}\mathcal{F}'_{jkn}.\end{aligned}$$

Let

$$\begin{aligned}\mathcal{M}_{ijn} &= \mathcal{E}_{iin}\mathcal{W}_{in}\mathcal{F}'_{jin} + \mathcal{F}_{ijn}\mathcal{W}_{jn}\mathcal{E}'_{jjn} \\ &\qquad + \Sigma^{*}v_{kn}\mathcal{F}_{ikn}\mathcal{F}'_{jkn}, \quad i \neq j, \qquad \text{<4.A.1.8>}\end{aligned}$$

$$Y_{ijn} = \Sigma^* F_{ijn} W_{kn} F'_{jkn}, \quad i \neq j: \qquad \text{<4.A.1.9>}$$

then,

$$U_{ijn} = v_{in} E_{iin} F'_{jin} + v_{jn} F_{ijn} E'_{jjn} + n^{-\frac{1}{2}} M_{ijn} + n^{-1} Y_{ijn},$$

which, for i, j $\neq$ **t** is equation <4.5.1.12>. Putting i = **t**, j $\neq$ **t** produces equation <4.5.1.13>, whilst doing the same for i $\neq$ **t**, j = **t** produces equation <4.5.1.14>.

<u>4.A.2.</u> In subsection 4.5.2., it is shown, by application of Rubin's Theorem, that if

$$U_n \xrightarrow{d} U,$$

then

$$E_{iin} \xrightarrow{d} E_{ii}, \quad i = 1,\ldots,t,$$

$$F_{ijn} \xrightarrow{d} F_{ij}, \quad i,j = 1,\ldots,t, \quad i \neq j,$$

and

$$W_{in} \xrightarrow{d} W_i, \quad i = 1,\ldots,t.$$

To establish the truth of the limiting equations <4.5.1.15>–<4.5.1.19>, it is necessary to show that

$$n^{-\frac{1}{2}} L_{ijn}, \; n^{-1} Y_{ijn}, \; n^{-\frac{1}{2}} M_{ijn}$$

all converge to zero, as $n \to \infty$. Examining the definitions of these matrices in equations <4.A.1.5>–<4.A.1.9>, it is clear that L_{ijn} and M_{ijn} (for i, j = 1,...**t**) are finite sums of products of matrices which either converge in distribution (F_{ijn} and W_{kn}) or converge to a constant ($v_{kn} I_{h_k}$) and hence themselves converge in distribution. Dividing by $n^{-\frac{1}{2}}$ then ensures that L_{ijn} and M_{ijn} converge in distribution to zero, and hence do so in probability. For Y_{ijn} this follows by a similar argument.

Chapter 5: Tests of Overidentifying Restrictions.

5.1. Introduction

5.1.1. The main purpose of this Chapter is to apply the statistical tests derived in Chapter 2 for a general constrained maximum likelihood problem to the linear simultaneous equations model, using the various limiting distribution results obtained in Chapter 3. In doing so, the intention is to examine the nature of the various test statistics, and look for simple or interesting ways to compute them.

It is useful to recall the notation for the general framework described in more detail in section 2.1.: the independent random m-vectors $y_1, \ldots, y_m$ have log-likelihood

$n^{-1} l_n(y; \theta)$,

and under the alternative hypothesis, equation <2.1.1.1>,

$$H_1: \quad \theta = \phi(\beta), \qquad \text{<5.1.1.1>}$$

θ being $s_0 \times 1$ and β being $r_1 \times 1$. For the purposes of deriving estimators, it is convenient to replace θ with ϕ. The null hypothesis adds the restrictions

$\beta = \lambda(\alpha)$,

where α is $r_0 \times 1$, so that the null hypothesis may be described by equation <2.1.1.2>,

$$H_0: \quad \theta = \phi(\beta), \quad \beta = \lambda(\alpha) \qquad \text{<5.1.1.2>}$$

or by equation <2.1.1.3>,

$$H_0: \quad \theta = \phi(\lambda(\alpha)) = \theta(\alpha); \qquad \text{<5.1.1.3>}$$

convenience determines which of these is used in a particular situation.

The maximum likelihood estimators of ϕ and β from the alternative hypothesis model are denoted

$\tilde{\phi}$, $\tilde{\beta}$,

whilst the estimators of θ, β and α from the null hypothesis model are denoted

$\tilde{\theta}$, $\tilde{\beta}_0$, $\tilde{\alpha}$.

The test statistics for testing <5.1.1.2> or <5.1.1.3> against <5.1.1.1> described in Chapter 2 - the Likelihood Ratio, Lagrange Multiplier, C-alpha, Wald minimum chi-squared - all have limit χ^2-distributions under the null hypothesis, with degrees of freedom equal to $r_1 - r_0$, the difference in the dimensionality of β and α. Using the null hypothesis in the form <5.1.1.2>, one can conclude directly that it is the $r_1 - r_0$ additional restrictions

$\beta = \lambda(\alpha)$

that are being subjected to test. However, it may occur that the dimensionality of θ, s_0, is equal to that of β, and then the alternative hypothesis model is described as just-identified, given that

$\theta = \phi(\beta)$

is invertible. It was noted in subsections 2.8.2., 2.9.3., 2.10.2. and 2.11.2. that when this occurs, the specific nature of the alternative hypothesis has no effect on the value of the Likelihood Ratio, Lagrange Multiplier, Wald and

C-alpha statistics, and thus one may interpret the test of <5.1.1.3> as being against any just-identified alternative. If this latter statement is interpreted as meaning "any choice of function ϕ and parameter β such that

$\theta = \phi(\beta)$

and

$\beta = \lambda(\alpha)$

implies that

$\theta = \theta(\alpha)$",

then it is not clear what restrictions are being tested, since the nature of the additional restrictions imposed on the just-identified model will change as different just-identified models are selected.

This ambivalence is present very strongly in the simultaneous equations literature of "tests of overidentifying restrictions": before proceeding to consider this literature, the definitions of the competing linear simultaneous equations models under the null and alternative hypotheses will be stated, and their parametric relationship with the general model noted.

5.1.2. From subsection 3.1.2., under the alternative hypothesis, the m-dimensional random vectors $y_1,\ldots,y_n$ are generated by the "reduced form" model,

$y_t = \Pi_1' x_t + v_{1t}, \quad t = 1,\ldots,n,$

or in observation matrix form,

$$Y = X\Pi_1 + V_1, \qquad \langle 5.1.2.1\rangle$$

where Π_1 is $k_1 \times m$, Y and V_1 $n \times m$. The matrix Π_1 is generated from the "structural form"

$$A_1'y_t + B_1'x_t = u_{1t}, \quad t = 1,\ldots,n,$$

or in observation matrix form,

$$YA_1 + XB_1 = U_1, \qquad \langle 5.1.2.2 \rangle$$

A_1 being $m \times m$ and B_1 $k_1 \times m$. The covariance matrices Ω_1, Σ_1 of the independent, zero mean random vectors v_{1t}, u_{1t} are linked by

$$\Sigma_1 = A_1'\Omega_1 A_1.$$

Defining

$$C_1' = (A_1' : B_1'), \quad Z_1 = (Y : X),$$

one can write

$$Z_1C_1 = U_1,$$

or

$$(I_m \otimes Z_1)\text{vec } C_1 = \text{vec } U_1,$$

or

$$(I_m \otimes Z_1)g_1 = u_1. \qquad \langle 5.1.2.3 \rangle$$

The parameter vector g_1 is restricted by

$$g_1 = K\delta + k, \qquad \langle 5.1.2.4 \rangle$$

where δ is $q_1 \times 1$, K is a known $m(m+k_1) \times q_1$ matrix of full column rank, and k a known $m(m+k_1) \times 1$ vector. In embedding this model into the general form, the vector ϕ is

$$\phi = \begin{bmatrix} v(\Omega_1) \\ \text{vec } \Pi_1 \end{bmatrix}, \qquad \langle 5.1.2.5 \rangle$$

whilst

$$\beta = \begin{bmatrix} v(\Omega_1) \\ \delta \end{bmatrix}. \qquad \text{<5.1.2.6>}$$

The null hypothesis model can be obtained, as in subsection 3.1.3., either by imposing the additional restrictions

$$\delta = L\gamma + r, \qquad \text{<5.1.2.7>}$$

where L is a known $q_1 \times q_0$ matrix of full column rank, γ a $q_0 \times 1$ vector of free parameters, and r a known $q_0 \times 1$ vector, or by writing out a full simultaneous equations model. The reduced form is

$$y_t = \Pi_0' x_t + v_{0t}, \quad t = 1,\ldots,n,$$

or

$$Y = X\Pi_0 + V_0, \qquad \text{<5.1.2.8>}$$

whilst the structural form is

$$A_0' y_t + B_0' x_t = u_{0t}, \quad t = 1,\ldots,n,$$

or

$$YA_0 + XB_0 = U_0.$$

Let

$$C_0' = (A_0' : B_0'),$$

$$g_0 = \text{vec } C_0, \quad u_0 = \text{vec } U_0;$$

then, the structural form is written in long vector form as

$$(I_m \otimes Z_1)g_0 = u_0,$$

and g_0 is restricted by

$$g_0 = H\gamma + h, \qquad \text{<5.1.2.9>}$$

H a known $m(m+k_1) \times q_0$ matrix of full column rank, h a known $m(m+k_1) \times 1$ vector. As observed in subsection 3.1.4., when the null hypothesis is true,

$$g_0 = H\gamma + h$$

$= KL\gamma + (Kr + k)$:

see equations <3.1.4.1> and <3.1.4.2>. The covariance matrices Ω_0, Σ_0 of v_{0t}, u_{0t} are linked by

$\Sigma_0 = A_0'\Omega_0 A_0$.

Embedding this model in the general model sets

$$\theta = \begin{bmatrix} v(\Omega_0) \\ \text{vec } \Pi_0 \end{bmatrix}, \qquad \text{<5.1.2.10>}$$

$$\alpha = \begin{bmatrix} v(\Omega_0) \\ \gamma \end{bmatrix}, \qquad \text{<5.1.2.11>}$$

and the relationship

$\beta = \lambda(\alpha)$

is just

$$\begin{bmatrix} v(\Omega_1) \\ \delta \end{bmatrix} = \begin{bmatrix} I_{\frac{1}{2}m(m+1)} & : & 0 \\ 0 & : & L \end{bmatrix} \begin{bmatrix} v(\Omega_0) \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 \\ r \end{bmatrix}.$$

<u>5.1.3.</u> To return to the question posed in subsection 5.1.1., of what is being tested by certain of the test statistics of Chapter 2, in the econometric literature the phrase "tests of over-identifying restrictions" means a test of the truth of equation <5.1.2.9> against any just-identified form of equation <5.1.2.4>. Which particular restrictions - that is, which choice of L and r in <5.1.2.7> is apparently of no interest; one might argue that a "misspecification test" is being considered. A specific choice of L and r in <5.1.2.7> then corresponds to a specification test: in fact, when <5.1.2.4> is overidentified, a specific choice of <5.1.2.7> is

automatically implied. These points are re-eexamined in what follows: they are mentioned now simply because an appreciation of the literature is thereby made easier.

5.1.4. The plan of the Chapter is as follows: in the next section, a critical survey of the literature is undertaken; typically this literature has considered only the "usual special case" of within equation, exclusion restrictions and unit normalisation rules, and thus, to present the arguments, some local notation (consistent with the general scheme) is required. Following this, the various test statistics outlined in Chapter 2 are applied, and compared: this is the substantive part of the Chapter, and it is believed that some of the test statistics have not yet appeared in the literature in the specific forms presented. The final part of the Chapter considers inference on a single over-identified equation by means of the LIML estimator discussed in section 3.7., and some estimators which are asymptotically equivalent to LIML under the null hypothesis.

5.2. Tests of Overidentifying Restrictions in the Simultaneous Equations Model: A Survey

5.2.1. The literature on this subject can be broken down into three general themes: tests based on the use of the LIML estimator; similar tests which use other limited information estimators like two-stage least squares or k-class, and tests based on the Likelihood Ratio, Wald or Lagrange Multiplier test principles, and which follow from the full information maximum likelihood estimation of the simultaneous equations model. Broadly speaking, this breakdown is also chronological. Although most of the papers which fall into the first category are well known, it is perhaps useful to review them, since at various points in the time, a certain amount of confusion has occurred in the literature.

Much of the discussion in the literature concerns the estimation of a single structural equation from the model <5.1.2.2>, say the first. This equation will be expressed as

$$\mathbf{y}_1 + \mathbf{Y}_1\mathbf{a}_1 + \mathbf{Y}_2\mathbf{a}_2 + \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 = u_{1.1}, \qquad \langle 5.2.1.1\rangle$$

where

$$Y = (\mathbf{y}_1 : \mathbf{Y}_1 : \mathbf{Y}_2): \quad n \times (1 + \mathbf{m}_1 + \mathbf{m}_2), \qquad \langle 5.2.1.2\rangle$$

$$X = (\mathbf{X}_1 : \mathbf{X}_2): \quad n \times (\mathbf{k}_{.1}, \mathbf{k}_{.2}). \qquad \langle 5.2.1.3\rangle$$

The equation is identified by the restrictions

$\mathbf{a}_2 = 0, \quad \mathbf{b}_2 = 0.$

In this literature, the rank condition for identification of $\mathbf{a}_1$ and $\mathbf{b}_1$ is expressed directly in terms of the first column of the relation

$\Pi_1 A_1 + B_1 = 0;$

partitioning this up to match the partition of Y and X in equations <5.2.1.2> and <5.2.1.3> produces

$$\begin{bmatrix} \pi_{1.10} & : & \Pi_{1.11} & : & \Pi_{1.12} \\ \pi_{1.20} & : & \Pi_{1.21} & : & \Pi_{1.22} \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ 0 \end{bmatrix} = 0, \qquad <5.2.1.4>$$

or writing

$\Pi^*_{1.11} = [\pi_{1.11} : \Pi_{1.11}];$

$\Pi^*_{1.21} = [\pi_{1.20} : \Pi_{1.21}],$

$a'_0 = (1 : a'_1),$

$$\begin{bmatrix} \Pi^*_{1.11} & : & \Pi_{1.12} \\ \Pi^*_{1.21} & : & \Pi_{1.22} \end{bmatrix} \begin{bmatrix} a_0 \\ 0 \end{bmatrix} + \begin{bmatrix} b_1 \\ 0 \end{bmatrix} = 0.$$

The rank condition is then that

$\text{rank } \Pi^*_{1.21} = m_1 = \text{rank } \Pi_{1.21},$

whilst the order condition is

$k_{.2} \geq m_1.$

5.2.2. Anderson and Rubin [1949, 1950] propose as the basis of a test of the null hypothesis that $k_{.2} - m_1$ of the elements of the vector b_2 are actually zero, the smallest root $1 + \tilde{v}$ of the determinental equation

$$\det[Y'_0(I_n - P_1)Y_0 - (1 + \tilde{v})Y'_0(I_n - P_X)Y_0] = 0 \qquad <5.2.2.1>$$

where

$Y_0 = (y_1 : Y_1),\ P_1 = X_1(X'_1X_1)^{-1}X'_1,$

$P_X = X(X'X)^{-1}X',$

by arguing that the null hypothesis on b_2 is equivalent to the null hypothesis that rank $\Pi^*_{1.21} = m_1$,

whilst the alternative hypothesis that the specified elements of $\mathbf{b}_2$ are not zero is equivalent to

rank $\Pi^*_{1.21} = m_1 + 1$.

The test statistic is the likelihood ratio test statistic,

$$- 2[-\tfrac{1}{2}n\log (1 + \tilde{v})] = n\log (1 + \tilde{v}) \overset{a}{\sim} \chi^2_{k_{.2}-m_1}, \qquad \langle 5.2.2.2\rangle$$

when the null hypothesis is true.

In contrast, Koopmans and Hood [1953, section 8] propose the same test statistic as a test of the null hypothesis

$$\mathbf{a}_2 = 0, \quad \mathbf{b}_2 = 0, \qquad \langle 5.2.2.3\rangle$$

with the alternative that at least one of the elementwise equalities fails, and argue that the null hypothesis is equivalent to the hypothesis

$$\text{rank } \Pi^*_{1.21} \leq m_1, \qquad \langle 5.2.2.4\rangle$$

whilst the alternative is equivalent to

rank $\Pi^*_{1.21} = m_1 + 1$.

The apparent conflict between the two null hypotheses, both when expressed in terms of the structural parameters and in terms of the rank of $\Pi^*_{1.21}$, has been the source of some confusion in the literature. For example, the rank of $\Pi^*_{1.21}$ on the null hypothesis <5.2.2.4> seems to allow the possibility that the first equation is unidentified; this observation led Savin [1975] to conclude, on the basis of results given by Anderson [1951], that

$$n\log (1 + \tilde{v})$$

cannot have the asserted limit χ^2-distribution under the null hypothesis <5.2.2.4>, since its distribution depends on the actual rank of $\Pi^*_{1.21}$. The overall conclusion reached by Savin [1975] is that the significance level of the Koopmans and Hood test is unknown. Kadane and Anderson [1977] did nothing to dispel this problem, since they proved that the null hypothesis <5.2.2.4> is equivalent to the statement that there exists at least one structure satisfying <5.2.2.3>.

An "independent observer" might suggest a resolution of the difficulty by demanding that the alternative hypothesis model given in equation <5.2.1.1> be identified when the null hypothesis is true; on the other hand, the discussion of the next chapter will show that some tests of restrictions are possible and feasible even in an unidentified model or equation.

In replying to a paper by Liu and Breen [1969], which amongst other things, misconstrued some of the points mentioned above, Fisher and Kadane [1972] set out what is perhaps the current conventional view of the LIML "single root test". For a model comprising a single equation like equation <5.2.1.1>, together with suitable exclusion restrictions, and the remaining equations of the model being in reduced form, the maximised log-likelihood function is equal to

$$-\tfrac{1}{2}n\log(1 + \tilde{v})$$

apart from an additive constant, $1 + \tilde{v}$ being the smallest root of an appropriate determinental equation similar to <5.2.2.1>. If the null hypothesis model is over-identified, and the alternative hypothesis being any just-identified model, the Likelihood Ratio statistic is given by equation <5.2.2.2>. Fisher and Kadane also observe that the alternative hypothesis model may include some unidentified structures, in the sense that they are observationally equivalent to some just-identified models. The validity of such a statement will be re-examined in the next chapter. Arguments concerning the consistency of this type of test are also given by Fisher and Kadane.

These ideas are extended more formally by Kadane [1974] to include cases where the first equation of the model is overidentified under both the null and alternative hypotheses, by means of exclusion restrictions: if $(1 + \tilde{v}_0)$ and $(1 + \tilde{v}_1)$ are the smallest roots of the appropriate determinental equations under the null and alternative hypotheses respectively, then the Likelihood Ratio statistic is

$$n\{\log(1 + \tilde{v}_0) - \log(1 + \tilde{v}_1)\}; \qquad \text{<5.2.2.5>}$$

when the alternative hypothesis is a just-identified equation, $\tilde{v}_1 = 0$. Kadane also formally examines the consistency of the test.

<u>5.2.3.</u> Using the least variance ratio principle of Koopmans and Hood [1953], one can write

$$1 + \tilde{v} = [\tilde{a}_0' Y_0' (I_n - P_X) Y_0 \tilde{a}_0]^{-1} \tilde{a}_0' Y_0' (I_n - P_1) Y_0 \tilde{a}_0$$

$\tilde{a}_0$ being the LIML estimator of a_0; Basmann [1960] suggested replacing $\tilde{a}_0$ by the two-stage least squares (or "generalised classical linear") estimator

$$\hat{a}_0' = (1 : \hat{a}_1'),$$

where

$$\hat{a}_1 = [Y_1'(P_X - P_1)Y_1]^{-1}Y_1'(P_X - P_1)y_1.$$

Defining

$$n\hat{v} = n[\hat{a}_0'Y_0'(I_n - P_X)Y_0\hat{a}_0]^{-1}\hat{a}_0'Y_0'(P_X - P_1)Y_0\hat{a}_0, \qquad \langle 5.2.3.1 \rangle$$

Basmann [1960] suggested the use of this as a test statistic parallel to that of equation <5.2.2.2>, and showed that

$$n\hat{v} \overset{a}{\sim} \chi^2_{k_{.2}-m_1}$$

under the null hypothesis.

Later, Kadane [1974] suggested the use of any k-class estimator of a_1 in equation <5.2.3.1>, and also developed an analogous test for the case where both the null and alternative hypotheses correspond to the first equation being over-identified. If $n\hat{v}_0$ and $n\hat{v}_1$ are the appropriate versions of equation <5.2.3.1> under the null and alternative hypotheses, the test statistic amounts to

$$n(\hat{v}_0 - \hat{v}_1) \overset{a}{\sim} \chi^2_t,$$

t being the additional number of exclusion restrictions imposed under the null hypothesis.

One can recognise the denominator of equation <5.2.3.1> as being proportional to a consistent estimator of σ^2, the disturbance variance of the first equation, and the numerator as the residual squared norm of the two-stage least squares estimator,

$$RSN = (y_1 + Y_1\hat{a}_1 + X_1\hat{b}_1)'P_X(y_1 + Y_1\hat{a}_1 + X_1\hat{b}_1):$$

this corresponds to the situation described in subsection 1.6.6., but where the matrix defining the metric is only positive semi-definite. The value of RSN here is zero if the first equation is just-identified (on the alternative hypothesis), so that one can regard the statistic $n\hat{v}$ of equation <5.2.3.1> as being equal to

$$\hat{\sigma}^{-2}(\mathrm{RSN}(\mathbf{H}_0) - \mathrm{RSN}(\mathbf{H}_1)). \qquad \text{<5.2.3.2>}$$

This viewpoint leads to the "asymptotic tests" discussed by Maddala [1974] and Hatanaka [1977]; in the identified equation

$$\mathbf{y}_1 + \mathbf{Y}_1\mathbf{a}_1 + \mathbf{X}_1\mathbf{b}_1 = \mathrm{u}_1, \qquad \text{<5.2.3.3>}$$

impose the $\mathbf{p}$ additional linearly independent restrictions

$$\mathbf{H}_{11}\mathbf{c}_1 = \mathbf{h}_1,$$

where

$$\mathbf{c}_1' = (\mathbf{a}_1' : \mathbf{b}_1'),$$

to form the null hypothesis model. A test of the truth of these restrictions is obtained from the limit distribution of an estimator $\mathbf{c}_1^*$ such that

$$n^{\frac{1}{2}}(\mathbf{c}_1^* - \mathbf{c}_1^0) \overset{a}{\approx} N(0, \Psi(\mathbf{c}_1^*; \mathbf{c}_1^0)),$$

where $\mathbf{c}_1^0$ is the true value of $\mathbf{c}_1$ under the null hypothesis; let an estimator of this covariance matrix be

$$\hat{\Psi}_n(\mathbf{c}_1^*) = \sigma_{11}^* N^*,$$

say, where σ_{11}^* is computed from the residual vector

$$\mathrm{u}_1^* = \mathbf{y}_1 + \mathbf{Y}_1\mathbf{a}_1^* + \mathbf{X}_1\mathbf{b}_1^*.$$

The test statistic is

$$\sigma_{11}^{*-1} n(\mathbf{H}_{11}\mathbf{c}_1^* - \mathbf{h}_1)'(\mathbf{H}_{11}N^*\mathbf{H}_{11}')^{-1}(\mathbf{H}_{11}\mathbf{c}_1^* - \mathbf{h}_1) \overset{a}{\approx} \chi_{\mathbf{p}}^2; \qquad \text{<5.2.3.4>}$$

when $\mathbf{c}_1^*$ is the two-stage least squares estimator $\hat{\mathbf{c}}_1$, this statistic can be expressed in the form of equation <5.2.3.2>.

This type of approach, and its variants, would seem to be well known, although not always explicitly discussed in the literature. Another approach, which exploits the nature of the two-stage least squares estimator, is due to Dhrymes [1969, 1970], with an extension to estimation by three-stage least squares by Morgan and Vandaele [1974], and produces limiting t or F-tests, in contrast to the limiting normal or χ^2-tests discussed so far. The argument works by replacing the denominator of equation <5.2.3.4> above by a quadratic form in the vector

$$n^{-\frac{1}{2}}X'u_1,$$

and which has a limiting χ^2-distribution, this limiting distribution being independent (in the limit) of the numerator quadratic form. The appropriate denominator quadratic form is

$$u_1'[P_X - P_X Z_1(Z_1'P_X Z_1)^{-1}Z_1'P_X]u_1 \overset{a}{\sim} \chi^2_{K_{.2}-m_1},$$

which is equal to the two-stage least squares residual squared norm: recall that

$$Z_1 = (Y_1 : X_1).$$

The matrix N^* which appears in equation <5.2.3.4> is defined by

$$N^* = (n^{-1}Z_1'P_X Z_1)^{-1}$$

in this two-stage least squares case.

Then, defining

$$w_n = np^{-1}\hat{\sigma}_{11}^{-1}(H_{11}\hat{c}_1 - h_1)'(H_{11}N^*H_{11}')^{-1}(H_{11}\hat{c}_1 - h_1),$$

$$x_n = (k_{.2} - m_1)^{-1}\hat{\sigma}_{11}^{-1}u_1'(P_X - P_X Z_1(Z_1'P_X Z_1)^{-1}Z_1'P_X)u_1,$$

one can claim that

$$w_n \xrightarrow{d} w, \quad x_n \xrightarrow{d} x,$$

and hence

$$(w_n)^{-1}x_n \xrightarrow{d} (w)^{-1}x \sim F(p, k_{.2} - m_1),$$

since the probability that $x = 0$ is zero, relative to the limiting distribution, for an over-identified equation. Maddala [1974] found, when doing a Monte Carlo experiment, that the Dhrymes test had relatively low power in some circumstances compared with the "asymptotic test".

Court [1974] gives a test, using three stage least squares estimation, which is very much in the spirit of the Koopmans and Hood approach: let each of the equations of the model be written in the form of equation <5.2.1.1>, as say

$$y_i = Y_{1i}a_{1i} + X_{1i}b_{1i} + Y_{2i}a_{2i} + X_{2i}b_{2i} = u_i, \quad i = 1,\dots,m,$$

where

$$[y_i \; \vdots \; Y_{1i} \; \vdots \; Y_{2i}]$$

is a permutation of the columns of the matrix Y of endogenous variables, and

$$[X_{1i} \; \vdots \; X_{2i}]$$

a permutation of the columns of the matrix X of exogenous variables. One can then define

$$Z_{1i} = (Y_{1i} \; \vdots \; X_{1i}), \quad Z_{2i} = (Y_{2i} \; \vdots \; X_{2i}),$$

$$c_{1i}' = (a_{1i}' \; \vdots \; b_{1i}'), \quad c_{2i}' = (a_{2i}' \; \vdots \; b_{2i}'),$$

and stack the m equations of the model in the form

$$y + Z_1c_1 + Z_2c_2 = u,$$

with

$\mathbf{y}' = (\mathbf{y}_1', \ldots, \mathbf{y}_m')$, $\quad \mathbb{Z}_i = \oplus_{j=1}^{m} \mathbb{Z}_{ij}$, $\quad i, j = 1,2$,

$\mathbf{c}_i' = (\mathbf{c}_{i1}', \ldots, \mathbf{c}_{im}')$, $i = 1,2$

$u' = (u_1, \ldots, u_m)$;

denote the covariance matrix of u by $\mathbb{S} \otimes I_n$. A test of the null hypothesis

$\mathbf{c}_2 = 0$

against the alternative

$\mathbf{c}_2 \neq 0$

is given by

$(\mathbf{y} - \mathbb{Z}_1\hat{\mathbf{c}}_1)'(\hat{\mathbb{S}} \otimes P_X)(\mathbf{y} - \mathbb{Z}_1\hat{\mathbf{c}}_1) \overset{a}{\approx} \chi^2_{\mathbf{p}}$,

under the null hypothesis, where $\hat{\mathbf{c}}_1$ and $\hat{\mathbb{S}}$ are the three-stage least squares estimators of $\mathbf{c}_1$ and $\mathbb{S}$ obtained under the null hypothesis; $\mathbf{p}$ here denotes the total number of overidentifying restrictions in the model. Court explains in a footnote

> "...it is not the hypothesis $\mathbf{c}_2 = 0$ which are being tested, because $\mathbf{c}_2$ cannot be estimated. (The statistic) tests whether certain estimable linear combinations of $\mathbf{c}_2$ are zero, these combinations being equal in number to the number of over-identifying restrictions specified in the model. If there are no over-identifying restrictions, there is no test."[Court [1974],p552]

Later, he states that the test statistic could be interpreted as testing the restrictions implied on the reduced form parameter matrix Π_1 by the overidentified structural specification.

<u>5.2.4.</u> Only in recent years have tests based on the use of

full information maximum likelihood estimation been developed, being those normally associated with the Likelihood Ratio, Wald and Lagrange Multiplier test principles. Hendry [1971] proposes a Likelihood Ratio test of over-identifying restrictions: under the null hypothesis, the simultaneous equations model is supposed to be over-identified, and (as will be shown in the next section) the maximised log-likelihood function corresponding to equation <3.3.4.1> is

$$n^{-1}l_n(y;\tilde{\theta}) = ms - \tfrac{1}{2}m - \tfrac{1}{2}\log\det\tilde{\Omega}_0,$$

using the restricted maximum likelihood estimators under the null hypothesis, whilst under any just-identified alternative hypothesis, the maximised log-likelihood is

$$n^{-1}l_n(y;\hat{\theta}) = ms - \tfrac{1}{2}m - \tfrac{1}{2}\log\det\hat{\Omega},$$

where the unrestricted, ordinary least squares, reduced form parameter estimators are used. Under the null hypothesis,

$$-2(l_n(y;\tilde{\theta}) - l_n(y;\hat{\theta})) \stackrel{a}{\sim} \chi^2$$

with degrees of freedom equal to the difference in the number of a priori restrictions on all the equations of the model, less m^2. If rejection occurs, the overidentifying restrictions are interpreted as being inconsistent with the sample information.

The work of Byron [1972] is explicitly an application of the arguments of Aitchison and Silvey [1958] and Silvey [1959] to the FIML estimation of a linear simultaneous equations model. In contrast to the approach taken in this thesis, Byron treats the vector g_0 of the "unrestricted" structural form parameters (see equation <3.1.2.5>) as the argument of the log-likelihood function (apart from the structural covariance matrix Σ_0) and imposes linear "constraint equation" restrictions of the form

$$\mathbf{R}g_0 = \mathbf{f} \qquad \langle 5.2.4.1\rangle$$

to express unit normalisation and exclusion restrictions. Observing that this parameterisation will produce a singular "information matrix", Byron suggests that the identifying restrictions may be introduced by "substitution". Thus, both the Lagrange Multiplier statistic and the Likelihood Ratio statistic for a test of the null hypothesis of equation <5.2.4.1> against any just-identified alternative will have, on the null hypothesis, a limit χ^2-distribution with degrees of freedom equal to the number of rows of $\mathbf{R}$ in <5.2.4.1> less the number of restrictions required for identification.

Wald, Lagrange Multiplier and Likelihood Ratio test statistics expressed in a form similar to the corresponding statistics described in Chapter 2, but appropriate for a constraint equation formulation, are given by Byron for the imposition of further zero restrictions in an already over-identified model, but he does not exploit the nature of the simultaneous equations model or the limit distribution of the FIML estimators to find interesting and useful expressions for these test statistics. Byron also suggests the use of individual Lagrange multipliers as test statistics for the individual restrictions being tested, and recognises the dangers inherent in such a procedure.

The Lagrange multiplier which appears in the Anderson and Rubin [1949] derivation of the LIML estimator is exploited by Byron to provide a test statistic corresponding to the Likelihood Ratio statistic given in equation <5.2.2.2>; as with the Koopmans and Hood test, there are some difficulties of interpretation.

In an important paper, Wegge [1978] proposes an estimator called "constrained indirect least squares", and which is very similar to three-stage least squares, except that the indirect least squares structural covariance matrix estimator is used instead of the two-stage least squares estimator typically used. The importance of this paper, from the point of view of this thesis, lies not so much in the development of the constrained indirect least squares estimator, but in the framework in which it is developed. Using a notation similar to that of Chapter 3, Wegge postulates a log-likelihood function

$l_n(y;\theta)$

where θ depends on a "structural" parameter vector, κ, say, and the dimensionality of κ exceeds that of θ. Constraint equation restrictions are then imposed,

$$\mathbf{e}_1 - \mathbf{g}_1(\kappa) = 0, \qquad \langle 5.2.4.2\rangle$$

$$\mathbf{e}_2 - \mathbf{g}_2(\kappa) = 0, \qquad \langle 5.2.4.3\rangle$$

(where $\mathbf{e}_1$, $\mathbf{e}_2$ are known vectors); the Lagrange multipliers associated with these restrictions are denoted $\mathbf{v}_1$, $\mathbf{v}_2$. The first set of these restrictions, equation <5.2.4.2>, is supposed to ensure that the parameter vector κ is locally identified. The maximum likelihood estimators of κ, v_1 and v_2 are found under the imposition of both <5.2.4.2> and <5.2.4.3>, and are denoted $\tilde{\kappa}$, $\tilde{v}_1$, $\tilde{v}_2$; when only <5.2.4.2> is imposed, the maximum likelihood estimators are denoted $\hat{\kappa}$, $\hat{v}_1$. By expressing the bordered information matrix of the former problem in terms of the bordered information matrix of the latter problem, Wegge is able to construct a "recursive" (i.e. two-step) estimator of κ, v_1 and v_2 satisfying both <5.2.4.2> and <5.2.4.3>, which will be denoted κ^*, v_1^*, v_2^*, in terms of the estimators $\hat{\kappa}$, $\hat{v}_1$ satisfying <5.2.4.2> only, and which has the same limiting distribution as the restricted maximum likelihood estimators $\tilde{\kappa}$, $\tilde{v}_1$, $\tilde{v}_2$. Unfortunately, the representation of κ^*, v_1^*, v_2^* is rather complex notationally, so that reference should be made to Wegge's paper for the algebraic detail.

To test the null hypothesis that both <5.2.4.2> and <5.2.4.3> hold, against the alternative that only <5.2.4.2> holds, Wegge proposes the Wald test statistic

$$n(\mathbf{e}_2 - \mathbf{g}_2(\hat{\kappa}))'[D_\kappa \mathbf{g}_2(\hat{\kappa})\Psi_n(\hat{\kappa})D_\kappa \mathbf{g}_2'(\hat{\kappa})]^{-1}(\mathbf{e}_2 - \mathbf{g}_2(\hat{\kappa})) \overset{a}{\sim} \chi^2_{\mathbf{p}},$$

where

$$n^{1/2}(\hat{\kappa} - \kappa) \overset{a}{\sim} N(0,\Psi(\hat{\kappa};\kappa)),$$

$\Psi_n(\hat{\kappa})$ is an estimate of $\Psi(\hat{\kappa};\kappa)$, and $\mathbf{p}$ the number of restrictions in <5.2.4.3>. Although this is proposed as a Wald test statistic, Wegge shows that it can be expressed in terms of the "two-step" Lagrange multiplier v_2^*.

In applying this test statistic to the simultaneous equations model, Wegge supposes that the restrictions in equation <5.2.4.2> are just-identifying, so that κ^* is the constrained indirect least squares estimator. A natural algebraic form for this estimator is not given by Wegge, but one was given in equation <3.6.3.2>. Wegge observes that testing the model with <5.2.4.2> and <5.2.4.3> imposed

> "...can only mean what subset of the overidentifying restrictions (<5.2.4.3>) can be retained" (Wegge, [1978],p442),

and that with each just-identified model (that is, with only <5.2.4.2> imposed) is associated a non-refutable true value of the parameter κ.

The other interesting aspect of Wegge's paper concerns a "multiple comparisons" approach to the detection of the subset of <5.2.4.2> which can be "retained" from a given just-identifying set of restrictions <5.2.4.2>; restrictions are selected for retention in order of the size of their contribution to the Wald statistic. That is, the restrictions which successively make the smallest contribution to the value of the statistic are retained. The subset selected for retention just before the test statistic becomes "significant" is then a "maximal retained subset": Wegge argues that this method is preferable to the Lagrange Multiplier approach of Byron [1972] described above.

The test statistic proposed by Wegge is clearly a structural parameter-based Wald test statistic; however, the underlying maximum likelihood framework might lead one to suppose that a test of all the over-identifying restrictions could be obtained using only the unrestricted reduced form parameter estimator: compare the minimum χ^2-principle described in section 2.10. . Byron [1974] attempts to produce a test of this type for a single structural equation like equation <5.2.3.3>, based on the existence of restrictions on the reduced form parameter matrix Π_0 if <5.2.3.3> is an over-identified equation.

To describe Byron's test statistic, consider the second equation in the system <5.2.1.4>. Partition up $\pi_{1.20}$ and $\Pi_{1.21}$ as

$$\pi'_{1.20} = (\pi^{*\prime}_{1.20} : \pi^{**\prime}_{1.20}), \quad \Pi'_{1.21} = (\Pi^{*\prime}_{1.21} : \Pi^{**\prime}_{1.21});$$

then, the equation

$$\pi_{1.20} + \Pi_{1.21}a_1 = 0$$

is equivalent to

$$\Pi^{*}_{1.21}a_1 + \pi^{*}_{1.20} = 0,$$
$$\Pi^{**}_{1.21}a_1 + \pi^{**}_{1.20} = 0.$$

Eliminating a_1 between the two equations yields a function of the elements of $\pi_{1.20}$ and $\Pi_{1.21}$,

$$\pi^{**}_{1.20} - \Pi^{**}_{1.21}(\Pi^{*}_{1.21})^{-1}\pi^{*}_{1.20} = 0,$$

and this expression, regarded as a function of the matrix Π_1, say $f(\Pi_1)$, forms the basis of Byron's Wald test. He finds the limit distribution of $f(\hat{\Pi})$ (i.e. using the least squares estimator of Π_1) under the null hypothesis that the over-identifying restrictions on equation <5.2.3.3> are true, using a Taylor series expansion method, and establishes a test statistic whose limit distribution is $\chi^2_{k_{.2}-m_1}$ (employing the "local" notation of subsection 5.2.1.). Byron argues that this test could be applied one equation at a time, and that the "double-counting" of restrictions on Π_1 that this entails may be circumvented; more importantly, he claims that the test statistic is invariant to the row partition of $\pi_{1.20}$ and $\Pi_{1.21}$ selected.

This assertion is disputed by Hwang [1980], who notes that each row partition is equivalent to the selection of a specific just-identified model out of the many that are valid when the over-identifying restrictions are false. Hwang goes on to show that the Byron test statistic is algebraically equivalent to the constrained indirect least squares Wald test statistic of Wegge [1978] for the case of a single over-identified equation.

Hwang also notes that Likelihood Ratio and Lagrange Multiplier tests (on the parameters of a single structural equation) are "symmetric" in the sense that one is not required to choose a specific just-identified model for the alternative hypothesis, whereas the Wald tests discussed in this section (but not those discussed so far in this thesis) are asymmetric in this sense. Whether this asymmetry is a problem depends on the investigator's objective: if tests of all the over-identifying restrictions are treated as misspecification tests (compare Hendry [1971]), lack of symmetry is a problem, since one does not obtain any information about the direction in which the misspecification has occurred.

One possible way around this problem is suggested by Hwang [1980]: arrange the restrictions imposed in a decreasing order according to the degree of a priori confidence placed on them, and simply regard the last $k_{.2} - m_1$ (in the single equation case discussed in subsection 5.2.1) as being the ones subject to test. This suggestion is very similar in spirit to Wegge's [1978] multiple comparison procedure.

5.2.5. One can see from the discussion of the preceding subsections the realisation that a "test of over-identifying restrictions" means a test of the restrictions placed on the reduced form parameters by an over-identified structural equation or model. Only when the alternative hypothesis model is over-identified is it clear what <u>structural</u> model rules if the null hypothesis is false, unless the investigator knows which of the possible non-refutable, just-identified structural models is true under the alternative hypothesis. The fact that in a situation where the alternative hypothesis model is just-identified, the "symmetric" and "asymmetric" test statistics are asymptotically equivalent (under the null hypothesis) only serves to obscure these issues even more.

5.3. Likelihood Ratio Tests of Over-Identifying Restrictions

5.3.1. Following on from the discussion in subsections 5.1.2. and 5.1.3., the intention is to provide a Likelihood Ratio test of the hypotheses

$$H_0: \quad \delta = L\gamma + r \qquad \langle 5.3.1.1\rangle$$

$$H_1: \quad \delta \neq L\gamma + r, \qquad \langle 5.3.1.2\rangle$$

in the linear simultaneous equations model described by equations <5.1.2.3> and <5.1.2.4>,

$$(I_m \otimes Z_1)g_1 = u_1,$$

$$g_1 = K\delta + k.$$

As observed in subsections 3.1.4. and 5.1.2., the simultaneous equations model ruling under the null hypothesis above can be defined as

$$(I_m \otimes Z_1)g_0 = u_0,$$

$$g_0 = H\gamma + h,$$

where, when the null hypothesis is true,

$$g_0 = K(L\gamma + r) + k$$

$$= KL\gamma + (Kr + k),$$

that is

$$H = KL, \quad h = Kr + k.$$

One can see from this the viewpoint expressed earlier in the thesis that the model under the null hypothesis is seen as a more restricted version of the model under the alternative hypothesis. However, it is also convenient (if only for estimation purposes) to express the hypotheses in the form

H_0': $g_0 = H\gamma + h$, <5.3.1.3>

H_1': $g_1 = K\delta + k$, <5.3.1.4>

leaving implicit the relationship

$\delta = L\gamma + r$.

To obtain a Likelihood Ratio statistic for testing <5.3.1.1> against <5.3.1.2>, the log-likelihood functions for the competing models are required. In section 3.3., maximum likelihood estimators for the null hypothesis parameters Π_0, $v(\Omega_0)$, γ (and δ) and for the alternative hypothesis parameters Π_1, $v(\Omega_1)$ and δ were found by maximising the appropriate log-likelihoods: the estimators were denoted $\tilde{\Pi}_0$, $v(\tilde{\Omega}_0)$, $\tilde{\gamma}$, $\tilde{\delta}_0$ and $\tilde{\Pi}_1$, $v(\tilde{\Omega}_1)$, $\tilde{\delta}$ respectively. The corresponding log-likelihood functions were given by equations <3.3.6.1> and <3.3.2.1> respectively; recall from subsection 5.1.2. that

$$\theta = \begin{bmatrix} v(\Omega_0) \\ \pi_0 \end{bmatrix}, \quad \phi = \begin{bmatrix} v(\Omega_1) \\ \pi_1 \end{bmatrix}.$$

The log-likelihoods are

$$n^{-1}l_n(y;\theta) = ms - \tfrac{1}{2}\log\det\Omega_0 - \tfrac{1}{2}n^{-1}\mathrm{tr}[\Omega_0^{-1}(Y - X\Pi_0)'(Y - X\Pi_0)]$$

and

$$n^{-1}l_n(y;\phi) = ms - \tfrac{1}{2}\log\det\Omega_1 - \tfrac{1}{2}n^{-1}\mathrm{tr}[\Omega_1^{-1}(Y - X\Pi_1)'(Y - X\Pi_1)]$$

using the reduced form models <5.1.2.8> and <5.1.2.1> respectively.

From equations <3.3.5.8> and <3.3.4.8>, one finds

$$\tilde{\Omega}_0 = n^{-1}(Y - X\tilde{\Pi}_0)'(Y - X\tilde{\Pi}_0),$$

$$\tilde{\Omega}_1 = n^{-1}(Y - X\tilde{\Pi}_1)'(Y - X\tilde{\Pi}_1),$$

and, for example,

$$\mathrm{tr}[\tilde{\Omega}_0^{-1}(Y - X\tilde{\Pi}_0)'(Y - X\tilde{\Pi}_0)] = n\,\mathrm{tr}\, I_m,$$

so that the maximised log-likelihood functions are

$$n^{-1}l_n(y;\tilde{\theta}) = ms - \tfrac{1}{2}\log\det\tilde{\Omega}_0 - \tfrac{1}{2}\mathrm{tr}\, I_m$$

$$= m(s - \tfrac{1}{2}) - \tfrac{1}{2}\log\det\tilde{\Omega}_0,$$

$$n^{-1}l_n(y;\tilde{\phi}) = m(s - \tfrac{1}{2}) - \tfrac{1}{2}\log\det\tilde{\Omega}_1. \qquad \text{<5.3.1.5>}$$

In these expressions, the maximum likelihood estimators of θ under the null and alternative hypotheses, $\tilde{\theta}$ and $\tilde{\phi}$ respectively, are used as convenient shorthands, consistent with the notation used in Chapter 2.

The Likelihood Ratio statistic of equations <2.8.1.1> is

$$LR = -2[l_n(y;\tilde{\theta}) - l_n(y;\tilde{\phi})]$$

which here equals

$$LR = n(\log\det\tilde{\Omega}_0 - \log\det\tilde{\Omega}_1), \qquad \text{<5.3.1.6>}$$

and, by section 2.8.1., has a limit χ^2-distribution with degrees of freedom equal to $r_1 - r_0$, the dimensionality of β and α respectively. In the simultaneous equations case, this is equal to

$$q_1 - q_0,$$

the difference in dimensionality of δ and γ.

When the alternative hypothesis <5.3.1.2> or <5.3.1.4> is just identified, the maximum likelihood estimators of Π_1 and Ω_1 are simply the least squares estimators $\hat{\Pi}$, $\hat{\Omega}$, the implied estimator of δ being the indirect least squares estimator; in this case,

$$mk_1 = q_1.$$

Then,

$$LR = n(\log\det\tilde{\Omega}_0 - \log\det\hat{\Omega}) \qquad \langle 5.3.1.7\rangle$$
$$\overset{a}{\sim} \chi^2_{mk_1-q_0},$$

which is the statistic given by Hendry [1971], and described in subsection 5.2.4., apart from the different notation to describe the degrees of freedom of the statistic.

One can see from this version of the Likelihood Ratio statistic that the same test statistic is obtained, whatever the nature of the just-identified alternative hypothesis model.

<u>5.3.2.</u> In subsection 2.8.1., the limiting χ^2-distribution of the Likelihood Ratio statistic was established by showing that it had the same limiting distribution as the random variable

$$n(\tilde{\theta} - \tilde{\phi})'\mathbf{I}(\theta^0)(\tilde{\theta} - \tilde{\phi}),$$

in which the limiting information matrix $\mathbf{I}(\theta^0)$ may be replaced by $\mathbf{I}_n(\tilde{\theta})$ or $\mathbf{I}_n(\tilde{\phi})$ to produce a test statistic which is clearly asymptotically equivalent to the Likelihood Ratio statistic.

It is quite interesting to investigate the nature of such statistics in the case of the simultaneous equations model: to do this, the finite sample information matrix of equation <3.4.2.2> (evaluated at the true value of θ under the null hypothesis, θ^0) will be used. One can then write

$$n(\tilde{\theta}-\tilde{\phi})'\mathbf{I}_n(\theta^0)(\tilde{\theta}-\tilde{\phi}) = \tfrac{1}{2}n v(\tilde{\Omega}_0 - \tilde{\Omega}_1)'\mathbf{D}_m(\Omega^{0^{-1}} \otimes \Omega^{0^{-1}})\mathbf{D}_m' v(\tilde{\Omega}_0 - \tilde{\Omega}_1)$$
$$+ n(\tilde{\pi}_0 - \tilde{\pi}_1)'(\Omega^{0^{-1}} \otimes n^{-1}X'X)(\tilde{\pi}_0 - \tilde{\pi}_1).$$

Next, one can show that the first term on the right hand side of this expression vanishes in probability, so that a statistic asymptotically equivalent to the Likelihood Ratio statistic can be obtained from the second term on the right hand side.

To show that

$n^{1/2}v(\tilde{\Omega}_0 - \tilde{\Omega}_1) \xrightarrow{P} 0,$

concentrate directly on

$n(\tilde{\Omega}_0 - \tilde{\Omega}_1) = n((\tilde{\Omega}_0 - \hat{\Omega}) - (\tilde{\Omega}_1 - \hat{\Omega})),$

$\hat{\Omega}$ being the least squares estimator of Π_0. Then,

$$\begin{aligned} n(\tilde{\Omega}_0 - \hat{\Omega}) &= (Y - X\tilde{\Pi}_0)'(Y - X\tilde{\Pi}_0) - (Y - X\hat{\Pi})'(Y - X\hat{\Pi}) \\ &= -Y'X(\tilde{\Pi}_0 - \hat{\Pi}) - (\tilde{\Pi}_0 - \hat{\Pi})'X'Y + \tilde{\Pi}_0'X'X\tilde{\Pi}_0 - \hat{\Pi}'X'X\hat{\Pi} \\ &= \tilde{\Pi}_0'X'X(\tilde{\Pi}_0 - \hat{\Pi}) - \hat{\Pi}'X'X(\tilde{\Pi}_0 - \hat{\Pi}), \end{aligned}$$

using the fact that

$X'Y = X'X\hat{\Pi}.$

Since

$\tilde{\Pi}_0,\ \hat{\Pi} \xrightarrow{P} \Pi^0,$

and

$n^{1/2}\mathrm{vec}\ (\tilde{\Pi}_0 - \hat{\Pi})$

has a proper limit normal distribution,

$n^{1/2}(\tilde{\Omega}_0 - \hat{\Omega}) \xrightarrow{P} (\Pi^{0\prime}M_X - \Pi^{0\prime}M_X)n^{1/2}(\tilde{\Pi}_0 - \hat{\Pi}) = 0.$

The same argument can be made for

$n^{1/2}(\tilde{\Omega}_1 - \hat{\Omega}),$

so that

$$LR \overset{a}{\sim} (\tilde{\pi}_0 - \tilde{\pi}_1)'(\Omega^{0-1} \otimes X'X)(\tilde{\pi}_0 - \tilde{\pi}_1). \qquad (5.3.2.1)$$

When the alternative hypothesis model is just-identified, $\tilde{\Pi}_1$

$= \hat{\Pi}$, and a further specialisation is obtained, for

$X(\tilde{\Pi}_0 - \hat{\Pi}) = \hat{V} - \tilde{V}_0$,

where $\tilde{V}_0$ and $\hat{V}$ are the reduced form residual matrices corresponding to the estimators $\tilde{\Pi}_0$, $\hat{\Pi}$:

$\tilde{V}_0 = Y - X\tilde{\Pi}_0, \quad \hat{V} = Y - X\hat{\Pi}$.

Using the result <1.6.1.2>, one can write

$$(\tilde{\pi}_0 - \hat{\pi})'(\Omega^{0-1} \otimes X'X)(\tilde{\pi}_0 - \hat{\pi}) = \mathrm{tr}[\Omega^{0-1}(\tilde{\Pi}_0 - \hat{\Pi})'X'X(\tilde{\Pi}_0 - \hat{\Pi})]$$
$$= \mathrm{tr}[\Omega^{0-1}(\hat{V} - \tilde{V}_0)'(\hat{V} - \tilde{V}_0)],$$

and

$$\hat{V}'\tilde{V}_0 = Y'(I_n - P_X)(Y - X\tilde{\Pi}_0)$$
$$= \hat{V}'\hat{V}.$$

One then obtains

$$\mathrm{tr}[\Omega^{0-1}(\hat{V} - \tilde{V}_0)'(\hat{V} - \tilde{V}_0)] = \mathrm{tr}[\Omega^{0-1}(\tilde{V}_0'\tilde{V}_0 - \hat{V}'\hat{V})]: \quad \text{<5.3.2.2>}$$

now making the choice of estimators of Ω^0 as $\tilde{\Omega}_0$ or $\hat{\Omega}$, in the case of $\tilde{\Omega}_0$, one has the test statistic

$$\mathrm{tr}[I_m - n(\tilde{V}_0'\tilde{V}_0)^{-1}\hat{V}'\hat{V}],$$

and in the case of $\hat{\Omega}$,

$$\mathrm{tr}[n(\hat{V}'\hat{V})^{-1}\tilde{V}_0'\tilde{V}_0 - I_m].$$

These statistics provide a link with the theory of multivariate linear hypotheses (see for example, Seber [1966]), in that they resemble Lawley's V and Hotelling's T^2-statistics. These results will turn out to be of interest in Chapter 9.

5.4. The Lagrange Multiplier Test Statistic

5.4.1. In Chapter 2, the Lagrange Multiplier test statistic for testing the hypotheses

H_0: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$,

H_1: $\theta = \phi(\beta)$

arose from the formal maximisation of

$n^{-1}l_n(y;\theta)$

subject to $\theta = \phi(\beta)$ and $\beta = \lambda(\alpha)$, with corresponding Lagrange multipliers ξ and ζ. The limiting distribution of the estimated Lagrange multipliers, $\tilde{\xi}$, $\tilde{\zeta}$ for the case of the simultaneous equations model was derived in subsection 3.4.3., where it was noted that because of the structure of θ, β and α, the covariance matrix of the limit normal distributions of $n^{1/2}\tilde{\xi}$ and $n^{1/2}\tilde{\zeta}$ were singular. In fact,

$$\Psi(\tilde{\xi};\psi_0^0) = \begin{bmatrix} 0 : & 0 \\ 0 : & \Psi(\tilde{\xi}_2;\psi_0^0) \end{bmatrix} = \begin{bmatrix} 0 : & 0 \\ 0 : & P'(\Omega^{0-1} \otimes M_x)P \end{bmatrix},$$

$$\Psi(\tilde{\zeta};\psi_0^0) = \begin{bmatrix} 0 : & 0 \\ 0 : & \Psi(\tilde{\zeta}_2;\psi_0^0) \end{bmatrix}$$

$$= \begin{bmatrix} 0 : & 0 \\ 0 : & K'(A^{0-1} \otimes B^{0\prime})P'(\Omega^{0-1} \otimes M_x)P(A^{0-1\prime} \otimes B^0)K \end{bmatrix},$$

where P is defined by equation <3.4.3.15>:

$$P = P(\alpha^0) = I_{mk_1} - (A^{0-1\prime} \otimes B^0)H[H'(\Sigma^{0-1} \otimes B^{0\prime}M_xB^0)H]^{-1}$$
$$\times H'(A^{0-1}\Omega^{0-1} \otimes B^{0\prime}M_x)$$

and A^0, B^0, Ω^0, Σ^0 are the true values of A_0, B_0, Ω_0, Σ_0 under the null hypothesis. In these expressions, ψ_0^0 collects together the true values under the null hypothesis as

$\psi_0^{0\prime} = (\theta^{0\prime}, \beta^{0\prime}, 0', \alpha^{0\prime}, 0')$:

see equation <3.4.3.4>.

The formal "score statistic" form of the Lagrange Multiplier statistic given in equation <2.9.2.1>,

$$LM = n^{-1}D_\theta l_n'(y;\tilde{\theta})\Phi(\tilde{\beta}_0)(\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi(\tilde{\beta}_0))^{-1}\Phi'(\tilde{\beta}_0)D_\theta l_n(y;\tilde{\theta}) \qquad <5.4.1.1>$$

is obtained on the assumption that the limit distribution of $n^{\frac{1}{2}}\tilde{\xi}$ is a nonsingular matrix; however, the argument in subsection 2.9.1. which shows that this statistic is asymptotically equivalent to the Likelihood Ratio statistic is unaffected by such singularities.

The matrix $\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi(\tilde{\beta}_0)$ appearing in equation <5.4.1.1> can be seen to have inverse matrix

$$(\Phi'(\tilde{\beta}_0)\mathbf{I}_n(\tilde{\theta})\Phi(\tilde{\beta}_0))^{-1} = \begin{bmatrix} 2\mathbf{L}_m\mathbf{S}_m(\tilde{\Omega}_0\otimes\tilde{\Omega}_0)\mathbf{S}_m\mathbf{L}_m' & 0 \\ 0 & n(K'(\tilde{\Sigma}_0^{-1}\otimes\tilde{B}_0'X'X\tilde{B}_0)K)^{-1} \end{bmatrix}$$

from equation <3.4.3.9>, whilst the score vector

$$n^{-1}D_\theta l_n(y;\tilde{\theta})$$

can be obtained from equations <3.4.2.1> or <3.3.3.1> as

$$n^{-1}D_\theta l_n(y;\tilde{\theta}) = \begin{bmatrix} -\tfrac{1}{2}\mathbf{D}_m(\tilde{\Omega}_0^{-1}\otimes\tilde{\Omega}_0^{-1})\mathbf{D}_m' v[\tilde{\Omega}_0 - n^{-1}(Y - X\tilde{\Pi}_0)'(Y - X\tilde{\Pi}_0)] \\ n^{-1}(\tilde{\Omega}_0^{-1}\otimes X')\text{vec }(Y - X\tilde{\Pi}_0) \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ n^{-1}(\tilde{\Omega}_0^{-1}\otimes X')\text{vec }\tilde{V}_0 \end{bmatrix}, \qquad <5.4.1.2>$$

since

$$\tilde{\Omega}_0 = n^{-1}(Y - X\tilde{\Pi}_0)'(Y - X\tilde{\Pi}_0) = \tilde{V}_0'\tilde{V}_0 :$$

see equation <3.3.5.8>. The matrix $\Phi(\tilde{\beta}_0)$ is, by equation <3.4.3.7>,

$$\Phi(\tilde{\beta}_0) = \begin{bmatrix} I_{\frac{1}{2}m(m+1)} & : & 0 \\ 0 & : & -(\tilde{A}_0^{-1\prime} \otimes \tilde{\Pi}_0)K \end{bmatrix},$$

so that, writing $\tilde{v}_0 = \text{vec}\, \tilde{V}_0$,

$$LM = \tilde{v}_0'(\tilde{\Omega}_0^{-1}\tilde{A}_0^{-1\prime} \otimes X\tilde{\Pi}_0)K(K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'X\tilde{\Pi}_0)K)^{-1}K'(\tilde{A}_0^{-1}\tilde{\Omega}_0^{-1} \otimes \tilde{\Pi}_0'X')\tilde{v}_0.$$

<5.4.1.3>

The relationship between the structural form and reduced form residual matrices $\tilde{U}_0$ and $\tilde{V}_0$ is

$$\tilde{V}_0 = \tilde{U}_0\tilde{A}_0^{-1},$$

so that, writing $\tilde{u}_0 = \text{vec}\, \tilde{U}_0$,

$$\tilde{v}_0 = (\tilde{A}_0^{-1} \otimes I_n)\tilde{u}_0;$$

recalling that

$$\tilde{\Sigma}_0 = \tilde{A}_0'\tilde{\Omega}_0\tilde{A}_0,$$

the Lagrange Multiplier statistic is

$$LM = \tilde{u}_0'(\tilde{\Sigma}_0^{-1} \otimes X\tilde{\Pi}_0)K(K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X'X\tilde{\Pi}_0)K)^{-1}K'(\tilde{\Sigma}_0^{-1} \otimes \tilde{\Pi}_0'X')\tilde{u}_0.$$

<5.4.1.4>

As in the general case discussed in subsection 2.9.2., the two statistics <5.4.1.3> and <5.4.1.4> can be regarded as the explained sum of squares (or "explained squared norm" - see subsection 1.6.6) in the regression of $\tilde{v}_0$ on $(\tilde{A}_0^{-1\prime} \otimes X\tilde{\Pi}_0)K$ in the metric of $(\tilde{\Omega}_0^{-1} \otimes I_n)$, or the regression of $\tilde{u}_0$ on $(I_m \otimes X\tilde{\Pi}_0)K$ in the metric of $(\tilde{\Sigma}_0^{-1} \otimes I_n)$.

<u>5.4.2.</u> When the alternative hypothesis model given by equations <5.3.1.2> or <5.3.1.4> is just identified, the matrix

$$(\tilde{A}_0^{-1} \otimes \tilde{\Pi}_0)K$$

is square and non-singular, so that the statistics <5.4.1.3>

and <5.4.1.4> collapse to

$$LM = \tilde{v}_o'(\tilde{\Omega}_o^{-1} \otimes P_X)\tilde{v}_o \qquad \text{<5.4.2.1>}$$

and

$$LM = \tilde{u}_o'(\tilde{\Sigma}_o^{-1} \otimes P_X)\tilde{u}_o, \qquad \text{<5.4.2.2>}$$

which are easily seen to be the explained squared norms of the regressions of $\tilde{u}_o$ and $\tilde{v}_o$ on $(I_m \otimes X)$ in the metric of $(\tilde{\Omega}_o^{-1} \otimes I_n)$ and $(\tilde{\Sigma}_o^{-1} \otimes I_n)$ respectively. However, consider writing <5.4.2.1> as a trace:

$$\begin{aligned} LM &= tr(\tilde{V}_o'P_X\tilde{V}_o\tilde{\Omega}_o^{-1}) \\ &= ntr(\tilde{V}_o'P_X\tilde{V}_o(\tilde{V}_o'\tilde{V}_o)^{-1}) \\ &= ntr((\tilde{V}_o'\tilde{V}_o)^{-1}\tilde{V}_o'P_X\tilde{V}_o). \end{aligned}$$

The matrix inside the trace is the regression coefficient matrix in the least squares regression of $P_X\tilde{V}_o$ on $\tilde{V}_o$, whilst $P_X\tilde{V}_o$ is itself the matrix of fitted values of the regression of $\tilde{V}_o$ on X. One can obtain exactly the same regression interpretation for $\tilde{U}_o$ using the statistic <5.4.2.2>. These results might be slightly more desirable than the "generalised least squares" regressions associated directly with equations <5.4.2.1> and <5.4.2.2>, simply because they only involve ordinary least squares regressions. It is also worth noting that the expression <5.4.2.2> is exactly the three-stage least squares residual squared norm if in fact $\tilde{\gamma}$ is the three-stage least squares estimator.

5.5. The C-alpha Test Statistic

5.5.1. This statistic was derived for the general problem in subsection 2.11.1, and is a by-product of the two-step estimation principle of section 2.6., specifically subsection 2.6.4., in that consistent but inefficient initial estimators α^*, $\beta^* = \lambda(\alpha^*)$, $\theta^* = \phi(\beta^*)$ with proper limit normal distributions are used to construct estimators $\hat{\alpha}$, $\hat{\beta} = \lambda(\hat{\alpha})$, $\hat{\theta} = \phi(\hat{\beta})$ with the same limit distribution as the maximum likelihood estimators $\tilde{\alpha}$, $\tilde{\beta}_0$, and $\tilde{\theta}$. In the process, Lagrange multipliers $\hat{\xi}$, $\hat{\zeta}$ are obtained, which depend on the vector

$$\mathbf{P}_n'(\alpha^*)D_\theta l_n(y;\theta^*), \qquad \langle 5.5.1.1\rangle$$

where $\mathbf{P}_n(\alpha^*)$ is defined by equation <2.6.2.4>,

$$\mathbf{P}_n(\alpha^*) = \mathbf{I}_{s_0} - \mathbf{B}(\alpha^*)(\mathbf{B}'(\alpha^*)\mathbf{I}_n(\theta^*)\mathbf{B}(\alpha^*))^{-1}\mathbf{B}'(\alpha^*)\mathbf{I}_n(\theta^*).$$

The C-alpha statistic then has the same general form as the Lagrange Multiplier statistic of equation <2.9.1.3>, except that the inefficient initial estimators α^*, β^*, θ^* are used, and the score vector $D_\theta l_n(y;\tilde{\theta})$ is replaced by equation <5.5.1.1> above,

$$CA = n^{-1}D_\theta l_n'(y;\theta^*)\mathbf{P}_n(\alpha^*)\Phi(\beta^*)[\Phi'(\beta^*)\mathbf{I}_n(\theta^*)\Phi(\beta^*)]^{-1}$$
$$\times\ \Phi'(\beta^*)\mathbf{P}_n'(\alpha^*)D_\theta l_n(y;\theta^*):$$

see equation <2.11.1.1>.

The quantities required to construct this statistic in the case of the simultaneous equations model have already been computed in the previous section; note that in subsection 3.5.1. it was assumed that

$\Omega_0^* = n^{-1}(Y - X\pi_0^*)'(Y - X\pi_0^*)$,

so that the first subvector in $D_\theta l_n(y;\theta^*)$ corresponding to $v(\Omega_0^*)$ vanishes, as in equation <5.4.1.2>. The structure of the projection matrix $\mathbf{P}_n(\alpha^*)$ can be deduced from equation <3.4.3.14>:

$$\mathbf{P}_n(\alpha^*) = \begin{bmatrix} 0 & : & 0 \\ 0 & : & P_n(\alpha^*) \end{bmatrix},$$

where $P_n(\alpha^*)$ in turn can be deduced from equation <3.4.3.15> as

$$P_n(\alpha^*) = I_{mk_1} - (A_0^{*-1\prime} \otimes D_0^*)H[H'(\Sigma_0^{*-1} \otimes D_0^{*\prime}X'XD_0^*)H]^{-1} \times H'(A_0^{*-1}\Omega_0^{*-1} \otimes D_0^{*\prime}X'X). \qquad \text{<5.5.1.2>}$$

Thus, for testing the hypotheses

$\mathbf{H}_0$: $\delta = L\gamma + r$,

$\mathbf{H}_1$: $\delta \neq L\gamma + r$,

the C-alpha statistic is

$$CA = v_0^{*\prime}(\Omega_0^{*-1} \otimes X)P_n(\alpha^*)(A_0^{*-1\prime} \otimes D_0^*)K[K'(\Sigma_0^{*-1} \otimes D_0^{*\prime}X'XD_0^*)K]^{-1} \times K'(A_0^{*-1} \otimes D_0^{*\prime})P_n'(\alpha^*)(\Omega_0^{*-1} \otimes X')v_0^*, \qquad \text{<5.5.1.3>}$$

where

$v_0^* = \text{vec } V_0^* = \text{vec}(Y - X\pi_0^*)$.

Let $P_{2n}(\alpha^*)$ be the projection matrix

$$P_{2n}(\alpha^*) = I_{mn}-(I_m \otimes XD_0^*)[H'(\Sigma_0^{*-1} \otimes D_0^{*\prime}X'XD_0^*)H]^{-1}H'(\Sigma_0^{*-1} \otimes D_0^{*\prime}X'): \qquad \text{<5.5.1.4>}$$

which satisfies

$$(\Sigma_0^{*-1} \otimes I_n)P_{2n}(\alpha^*) = P_{2n}'(\alpha^*)(\Sigma_0^{*-1} \otimes I_n): \qquad \text{<5.5.1.5>}$$

then,

$$(I_m \otimes X)P_n(\alpha^*)(A_0^{*-1\prime} \otimes D_0^*)K = (A_0^{*-1\prime} \otimes I_n)P_{2n}(\alpha^*)(I_m \otimes XD_0^*)K.$$

Together with the relationship

$$v_0^* = (A_0^{*-1\prime} \otimes I_n)u_0^*,$$

the C-alpha statistic of equation <5.5.1.3> can be written in terms of the structural form residual vector u_0^* as

$$CA = u_0^{*\prime}(\Sigma_0^{*-1} \otimes I_n)P_{2n}(\alpha^*)(I_m \otimes X\Pi_0^*)K[K'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)K]^{-1}$$
$$\times K'(I_m \otimes \Pi_0^{*\prime}X')P_{2n}'(\alpha^*)(\Sigma_0^{*-1} \otimes I_n)u_0^*$$
$$= u_0^{*\prime}P_{2n}'(\alpha^*)(\Sigma_0^{*-1} \otimes X\Pi_0^*)K[K(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)K]^{-1}$$
$$\times K'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')P_{2n}(\alpha^*)u_0^*. \qquad <5.5.1.6>$$

This statistic can be obtained as the explained squared norm in the regression of $P_{2n}(\alpha^*)u_0^*$ on $(I_m \otimes X\Pi_0^*)K$ with respect to the metric $(\Sigma_0^{*-1} \otimes I_n)$. If the vector $P_{2n}(\alpha^*)u_0^*$ is interpreted as a residual vector $u_0^{\dagger}$, say, then the statistic <5.5.1.6> has exactly the same form as the Lagrange Multiplier statistic of equation <5.4.1.4>.

<u>5.5.2.</u> When the alternative hypothesis corresponds to a just-identified model, the C-alpha statistic collapses, just as the Lagrange Multiplier does, to

$$CA = v_0^{*\prime}(\Omega_0^{*-1} \otimes X)P_n(\alpha^*)(\Omega_0^* \otimes (X'X)^{-1})P_n'(\alpha^*)(\Omega_0^{*-1} \otimes X')v_0^*,$$
$$<5.5.2.1>$$

or to

$$CA = u_0^{*\prime}P_{2n}'(\alpha^*)(\Sigma_0^{*-1} \otimes P_X)P_{2n}(\alpha^*)u_0^*, \qquad <5.5.2.2>$$

and can be given similar interpretations as explained squared norms in certain regressions. By noting that

$$(I_m \otimes X)v_0^* = (I_m \otimes X'X)(\hat{\pi} - \pi_0^*),$$

the version <5.5.2.1> can be written as

$$CA = (\hat{\pi} - \pi_0^*)'[(\Omega_0^{*-1} \otimes X'X) - (\Omega_0^{*-1}A_0^{*-1\prime} \otimes X'X\Pi_0^*)H$$

$$\times\ [H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^{*})H]^{-1}(A_0^{*-1}\Omega_0^{*-1} \otimes \Pi_0^{*\prime}X'X)](\hat{\pi} - \pi_0^{*}).$$

<5.5.2.3>

The first term in this expression is very similar to a version of the Likelihood Ratio statistic, equation <5.3.2.1>, whilst the second term shows the need for a "correction factor" arising from the use of an inefficient estimator. The statistic itself is the residual squared norm of the regression of $\hat{\pi} - \pi_0^{*}$ on $(A_0^{*-1\prime} \otimes \Pi_0^{*})H$, in the metric of $(\Omega_0^{*-1} \otimes X'X)$ or, equivalently, the regression of

$(I_m \otimes X)(\hat{\pi} - \pi_0^{*}) = (v_0^{*} - \hat{v})$

on $(A_0^{*-1\prime} \otimes X\Pi_0^{*})H$ in the metric of $(\Omega_0^{*-1} \otimes I_n)$. Since $\hat{v}$ and the matrix $(A_0^{*-1\prime} \otimes X\Pi_0^{*})H$ are orthogonal to each other,

$(I_m \otimes X')\hat{v} = 0,$

one can see that the regression coefficient vector in this regression is precisely the "update term" in the expression <3.5.1.2> for the two-step estimator $\hat{\hat{\gamma}}$ of γ. Again, one can show that the vector of regression coefficients in the regression producing the C-alpha statistic of equation <5.5.2.2> is the update term in the alternative expression <3.5.1.5> for the two-step estimator $\hat{\hat{\gamma}}$.

These results display a useful feature of two-step estimation: by a single multivariate regression, using inefficient initial estimators, one can obtain an estimator which is asymptotically equivalent to the maximum likelihood estimator $\tilde{\gamma}$, and simultaneously, a test statistic for the over-identifying restrictions of the model. In addition, estimation is required only under the null hypothesis, apart from the need to calculate the unrestricted least squares estimator $\hat{\pi}$.

5.6. Wald Test Statistics

5.6.1. The null hypothesis of equation <5.3.1.1>,

$\delta = L\gamma + r,$

can be expressed purely as constraint equations on the vector δ of the alternative hypothesis model of equations <5.1.2.3> and <5.1.2.4>,

$(I_m \otimes Z_1)g_1 = u_1,$

$g_1 = K\delta + k.$

Let $\mathbf{D}$ be a matrix with $\mathbf{q}_1 - \mathbf{q}_0$ linearly independent rows and $\mathbf{q}_1$ columns, such that

$N(\mathbf{D}) = \mathbf{C}(L):$

then, by the results of subsection 1.6.1.,

$\mathbf{D}\delta = \mathbf{D}r \Leftrightarrow \delta = L\gamma + r.$

From section 3.4., the maximum likelihood estimator of δ from the alternative hypothesis model, $\tilde{\delta}$, has a limit normal distribution with covariance matrix given by equation <3.4.3.10>:

$\Psi(\tilde{\delta};\psi_1^0) = [K'(\Sigma^{0-1} \otimes \Omega^{0\prime}M_X\Omega^0)K]^{-1},$

and ψ_1^0 is defined in equation <3.4.3.2>. Thus, it follows that

$n^{1/2}(\mathbf{D}\tilde{\delta} - \mathbf{D}r) \overset{a}{\approx} N(0,\mathbf{D}\Psi(\tilde{\delta};\psi_1^0)\mathbf{D}'),$

and a limit χ^2-statistic under the null hypothesis is given by

$$W = (\mathbf{D}\tilde{\delta} - \mathbf{D}r)'\{\mathbf{D}[K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)K]^{-1}\mathbf{D}'\}^{-1}(\mathbf{D}\tilde{\delta} - \mathbf{D}r) \overset{a}{\approx} \chi^2_{\mathbf{q}_1-\mathbf{q}_0}. \qquad \text{<5.6.1.1>}$$

Even when the alternative hypothesis is just-identified, so that $\tilde{\delta}$ becomes the indirect least squares estimator given in equation <3.6.3.1>, and $\tilde{\Omega}_1$ becomes the least squares estimator $\hat{\Omega}$, there is no particular simplification in the test statistic.

It is quite clear that this test statistic is "asymmetric" in the sense that the nature of the structure of the alternative hypothesis model must be known, even when the alternative is just-identified.

The Wald test statistics of section 2.10. are designed to be symmetric in the case of a just-identified alternative hypothesis model: the next subsection constructs a test of the hypothesis of equation <5.3.1.1>,

$\delta = L\gamma + r$,

for this case.

<u>5.6.2.</u> It was noted in subsection 3.6.1. that the criterion function of the minimum chi-squared estimator of the structural parameter α of the null hypothesis of equation <5.1.1.3>,

$\mathbf{H}_0$: $\theta = \theta(\alpha)$,

is

$(\hat{\theta} - \theta(\alpha))'\mathbf{I}_n(\hat{\theta})(\hat{\theta} - \theta(\alpha))$,

$\hat{\theta}$ being the unrestricted maximum likelihood estimator of θ; this criterion function collapses in the simultaneous equations model to equation <3.6.1.1>,

$$n^{-1}(\hat{\pi} - \theta_2(\gamma))'(\hat{\Omega}^{-1} \otimes X'X)(\hat{\pi} - \theta_2(\gamma)), \qquad \text{<5.6.2.1>}$$

where

$$\pi_0 = \theta_2(\gamma) \qquad \text{<5.6.2.2>}$$

expresses the composite function dependence of π_0 on γ through

$$\pi_0 = \text{vec}\ (B_0A_0^{-1})$$

as a function of g_0, and

$$g_0 = H\gamma + h.$$

Letting γ^* denote the formal minimiser of equation <5.6.2.1> above, the Wald test statistic for a test of

$$H_0: \quad \delta = L\gamma + r,$$

against

$$H_1: \quad \delta \neq L\gamma + r,$$

when the alternative hypothesis corresponds to a just-identified model, is

$$\begin{aligned} W &= (\hat{\pi} - \theta_2(\gamma^*))'(\hat{\Omega}^{-1} \otimes X'X)(\hat{\pi} - \theta_2(\gamma^*)) \\ &= (\hat{\pi} - \pi_0^*)'(\hat{\Omega}^{-1} \otimes X'X)(\hat{\pi} - \pi_0^*) \overset{a}{\approx} \chi^2_{mk_1 - q_0}, \end{aligned} \qquad \text{<5.6.2.3>}$$

where

$$\pi_0^* = \theta_2(\gamma^*).$$

This statistic may be interpreted as providing a test that

$$\hat{\pi} - \theta_2(\gamma^*) = 0,$$

or that $\hat{\pi}$ satisfies the restrictions

$$g(\hat{\pi}) = 0$$

contained in the equations of <5.6.2.2>: put slightly differently, a test of the restrictions imposed on the reduced form by the overidentifying restrictions. This remark may be compared with the comment given by Court [1974] stated at the end of subsection 5.2.3.; the test statistic <5.6.2.3> may be compared with Byron's [1974] test statistic described in subsection 5.2.4. .

Practical calculation of the statistic <5.6.2.3> will require iterative methods, and it might therefore be more straightforward to consider the corresponding test statistic based on the "linearised minimum chi-squared" estimator described in subsection 3.6.1. . This estimator, γ^{∇}, minimises

$$n^{-1}((\hat{\pi} - \pi_0^*) + (A_0^{*-1\prime} \otimes B_0^*)H(\gamma^{\nabla} - \gamma^*))'(\hat{\Omega}^{-1} \otimes X'X)$$
$$\times\ ((\hat{\pi} - \pi_0^*) + (A_0^{*-1\prime} \otimes B_0^*)H(\gamma^{\nabla} - \gamma^*)), \quad \text{<5.6.2.4>}$$

where π_0^*, A_0^*, and B_0^* are the estimators corresponding to the consistent but inefficient initial estimator γ^*. The general results show that n times equation <5.6.2.4> has the same limit χ^2-distribution as the W-statistic of equation <5.6.2.3>, and is clearly equal to the residual squared norm of the regression of $\hat{\pi} - \pi_0^*$ on $(A_0^{*-1\prime} \otimes B_0^*)H$ in the metric of $\hat{\Omega} \otimes X'X$, or equivalently, of

$$(I_m \otimes X)(\hat{\pi} - \pi_0^*) = (v_0^* - \hat{v})$$

on $(A_0^{*-1\prime} \otimes XB_0^*)H$ in the metric of $\hat{\Omega} \otimes I_n$. This is exactly the same statistic as the C-alpha statistic of equation <5.5.2.3>, apart from the use of $\hat{\Omega}$ instead of Ω_0^*.

It is clear that this linearised Wald statistic needs unrestricted estimation of the parameters π_0 and Ω_0, as well as inefficent but restricted estimation of the null hypothesis parameters. This latter requirement, which conflicts with the usual view of the Wald statistic requiring estimation only under the alternative hypothesis, arises solely from the need to linearise the minimum chi-squared criterion function.

5.6.3. The Wald test statistic for the simultaneous equations equivalent of the null hypothesis

H_0: $\theta = \phi(\beta)$, $\beta = \lambda(\alpha)$

with the overidentified alternative

H_1: $\theta = \phi(\beta)$

is very straightforward, since the equivalent of $\beta = \lambda(\alpha)$ is the equation

$\delta = L\gamma + r$.

The minimum chi-squared estimator of γ, denoted γ^*, comes formally from minimising <3.6.2.1>:

$$(\tilde{\delta} - L\gamma - r)'K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)K(\tilde{\delta} - L\gamma - r), \qquad \text{<5.6.3.1>}$$

but it is shown in subsection 3.6.2. that γ^* (defined in equation <3.6.2.3>) can be obtained directly by a regression of

$-(I_m \otimes X\tilde{\Omega}_1)h$ on $(I_m \otimes X\tilde{\Omega}_1)H$

in the metric of $\tilde{\Sigma}_1^{-1} \otimes I_n$. This turns on the fact, established in equation <3.6.2.2>, that

$K(\tilde{\delta} - r) = \tilde{g}_1 - h$.

Thus, using equations <3.6.2.2> and <3.6.2.3>,

$$\begin{aligned}(I_m \otimes X\tilde{\Omega}_1)K(\tilde{\delta} - L\gamma^* - r) &= (I_m \otimes X\tilde{\Omega}_1)(I_{m(m+k_1)} - H[H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)H]^{-1} \\ &\qquad \times H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1))K(\tilde{\delta} - r) \\ &= (I_m \otimes X\tilde{\Omega}_1)\tilde{g}_1 - (I_{mn} - (I_m \otimes X\tilde{\Omega}_1)H \\ &\qquad \times [H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'X\tilde{\Omega}_1)H]^{-1}H'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'))(I_m \otimes X\tilde{\Omega}_1)h \\ &= -P_{2n}(\tilde{\beta})(I_m \otimes X\tilde{\Omega}_1)h,\end{aligned}$$

since $(I_m \otimes \tilde{\Omega}_1)\tilde{g}_1 = 0$ and where $P_{2n}(.)$ is defined by equation <5.5.1.4>: note that $\tilde{\beta}$ corresponds to $\tilde{\delta}$ and hence to $\tilde{\Omega}_1$, $\tilde{\Sigma}_1$.

The residual squared norm of the regression of $-(I_m \otimes X\tilde{\Omega}_1)h$

on $(I_m \otimes X\tilde{\Pi}_1)H$ in the metric of $(\tilde{\Sigma}_1^{-1} \otimes I_n)$ is therefore the minimum value of the expression in equation <5.6.3.1>, and the Wald statistic is

$$W = n(H\gamma^* + h)'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Pi}_1'X'X\tilde{\Pi}_1)(H\gamma^* + h) \qquad \text{<5.6.3.2>}$$

$$= ng_0^{*\prime}(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Pi}_1'X'X\tilde{\Pi}_1)g_0^*, \qquad \text{<5.6.3.3>}$$

where

$$g_0^* = H\gamma^* + h.$$

Note that this cannot be expressed in terms of the corresponding residual vector

$$u_0^* = (I_m \otimes Z_1)g_0^*$$

because the generating regression uses "fitted values" $X\tilde{\Pi}_1$ rather than the actual values Z_1.

In the just-identified case, when $\tilde{\delta}$ is the indirect least squares estimator of equation <3.6.3.1>, the statistic <5.6.3.3> changes to the extent that $\tilde{\Pi}_1 = \hat{\Pi}$ and $\hat{\Pi}'X'X\hat{\Pi} = Z_1'P_XZ_1$:

then,

$$(I_m \otimes X\hat{\Pi})g_0^* = (I_m \otimes P_X)Z_1g_0^*$$

$$= (I_m \otimes P_X)u_0^*,$$

so that the statistic could be obtained by a regression of u_0^* on $I_m \otimes X$ in the metric of $\tilde{\Sigma}_1^{-1} \otimes I_n$, that is, by minimising

$$n(H\gamma^* + h)'(\tilde{\Sigma}_1^{-1} \otimes Z_1'P_XZ_1)(H\gamma^* + h)$$

with respect to γ^*. Since this produces the constrained indirect least squares estimator of equation <3.6.3.2>, this is the criterion function for that estimator; furthermore, replacing $\tilde{\Sigma}_1$ by the two-stage least squares estimator $\hat{\Sigma}_1$ say, produces the well-known three-stage least squares estimator.

Note that the statistic <5.6.3.3> is "asymmetric", even when the alternative hypothesis model is just identified.

5.6.4. It is also interesting to consider the relationship between the Wald statistics <5.6.1.1> and <5.6.3.3>: in general, they are not equal, but are "nearly equal". The nature of the similarity can be seen by using analogies from the theory of linear estimation. Let

$\mathbf{x} = (I_m \otimes X\tilde{\Omega}_1)K(\tilde{\delta} - r)$,

$\mathbf{F} = (I_m \otimes X\tilde{\Omega}_1)K$,

$\mathbf{G} = \tilde{\Sigma}_1^{-1} \otimes I_n$:

then, the statistic <5.6.3.3> is obtained by minimising

$$(\mathbf{x} - \mathbf{F}L\gamma)'\mathbf{G}(\mathbf{x} - \mathbf{F}L\gamma). \qquad \text{<5.6.4.1>}$$

Since

$\delta = L\gamma + r$

is equivalent to

$\mathbf{D}\delta = \mathbf{D}r$,

it follows from linear estimation theory that the minimum value of equation <5.6.4.1> is the minimum value of

$$(\mathbf{x} - \mathbf{F}\delta)'\mathbf{G}(\mathbf{x} - \mathbf{F}\delta) \qquad \text{<5.6.4.2>}$$

subject to

$\mathbf{D}\delta = \mathbf{D}r$.

The statistic <5.6.1.1> is then equal to the difference of the constrained minimum of <5.6.4.2> and its unconstrained minimum: only when this latter minimum value is identically zero will the two statistics <5.6.1.1> and <5.6.3.3> coincide. Essentially, the inequality stems from the fact that

$\tilde{\delta} = -[K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'Z_1)K]^{-1}K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Omega}_1'X'Z_1)k$

and not

$$\tilde{\delta} = -[K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Pi}_1'X'X\tilde{\Pi}_1)K]^{-1}K'(\tilde{\Sigma}_1^{-1} \otimes \tilde{\Pi}_1'X'X\tilde{\Pi}_1)k.$$

Estimators like three-stage least squares and constrained indirect least squares, which have the desired symmetry property, will give the equality between the two Wald statistics.

5.7. Difference Statistics

5.7.1. When estimation of both the null and alternative hypothesis models of equations <5.3.1.3> and <5.3.1.4>,

$$\mathbf{H}_0': \quad g_0 = H\gamma + h, \qquad \langle 5.7.1.1\rangle$$

$$\mathbf{H}_1': \quad g_1 = K\delta + k, \qquad \langle 5.7.1.2\rangle$$

is of interest, one can consider the possibility of testing the null hypothesis

$$\mathbf{H}_0: \quad \delta = L\gamma + r \qquad \langle 5.7.1.3\rangle$$

against

$$\mathbf{H}_1: \quad \delta \neq L\gamma + r \qquad \langle 5.7.1.4\rangle$$

by means of a test of $\mathbf{H}_0'$ against some just-identified alternative $\mathbf{H}_2'$, and also a test of $\mathbf{H}_1'$ against the same alternative $\mathbf{H}_2'$. Such a procedure will use the difference of the two test statistics.

The just-identified model of the hypothesis $\mathbf{H}_2'$ which will serve as alternative to the "null" hypotheses $\mathbf{H}_0'$, $\mathbf{H}_1'$ may be defined by

$$(I_m \otimes Z_1)g_2 = u_2,$$

with the covariance matrix of the vector u_2 being Σ_2, and the just-identifying restrictions expressed as the hypothesis

$$\mathbf{H}_2': \quad g_2 = \mathbf{J}\mathbf{c} + \mathbf{d}, \qquad \langle 5.7.1.5\rangle$$

where $\mathbf{J}$ is $m(m + k_1) \times q_2$, $\mathbf{c}$ $q_2 \times 1$. One can suppose that γ and $\mathbf{c}$, δ and $\mathbf{c}$ are connected by

$$\gamma = \mathbf{M}\mathbf{c} + \mathbf{f}, \qquad \langle 5.7.1.6\rangle$$

$$\delta = \mathbf{N}\mathbf{c} + \mathbf{h}. \qquad \langle 5.7.1.7\rangle$$

5.7.2. Of the various test statistics for the hypotheses <5.7.1.3> and <5.7.1.4> considered so far, one can distinguish two groups of statistics: those that decompose automatically into a test of <5.7.1.6>, and a test of <5.7.1.7> (with alternative hypotheses being defined by failure of the equality in these hypotheses), and those that decompose "asymptotically". In the first category falls the Likelihood Ratio statistic <5.3.1.6>, which can be written as

$$LR = n((\log\det\tilde{\Omega}_0 - \log\det\hat{\Omega}) - (\log\det\tilde{\Omega}_1 - \log\det\hat{\Omega})),$$

since, as noted in equation <5.3.1.7>, the first term is appropriate for testing the null hypothesis <5.7.1.3> against a just-identified alternative.

5.7.3. The general result of subsection 2.12.2. showed that the overall Lagrange Multiplier test statistic of

H_0: $\theta = \theta(\alpha)$

against

H_1: $\theta = \phi(\beta)$,

when both hypotheses correspond to over-identified models, is only asymptotically equivalent to the difference of the Lagrange Multiplier statistics LM_0 and LM_1 for testing each of these two "null" hypotheses against the hypothesis

H_2: θ unrestricted

(i.e. just-identified). This feature carries over to the case of the simultaneous equations model: from equation <5.4.2.2>, the appropriate Lagrange Multiplier statistics are clearly

$$LM_0 = \tilde{u}_0'(\tilde{\Sigma}_0^{-1} \otimes P_X)\tilde{u}_0,$$

and

$LM_1 = \tilde{u}_1'(\tilde{\Sigma}_1^{-1} \otimes P_X)\tilde{u}_1.$

Since the Lagrange Multiplier statistic in general only requires quantities estimated under the "null hypothesis", there are fortunately no notational complications. Only when the same covariance matrix estimator Σ_1^* and reduced form parameter matrix Ω_1^*, say, are used to find the estimators $\tilde{\gamma}$ and $\tilde{\delta}$ which underlie the residual vectors $\tilde{u}_0$, $\tilde{u}_1$ will the Lagrange Multiplier statistic of equation <5.4.1.4>, denoted LM, coincide with $LM_0 - LM_1$: this will occur, for example, when Σ_1^* is the two-stage or indirect least squares estimator, and Ω_1^* the least squares estimator $\hat{\Omega}$, for then, $\tilde{\gamma}$ and $\tilde{\delta}$ are three-stage or constrained indirect least squares estimators. Otherwise, one has to rely on an asymptotic equivalence:

$LM \overset{a}{\approx} LM_0 - LM_1.$

<u>5.7.4.</u> Similar results hold for the C-alpha statistics of subsection 5.5.2.: suppose that the same initial estimators

γ^*, g_0^*, Σ_0^*, Ω_0^* and

$\delta_0^* = L\gamma_0^* + r$

are used to construct two-step estimators for the models defined by equations <5.7.1.1> an <5.7.1.2>. Under this assumption, the same residual vector u_0^* and covariance matrix Σ_0^* is used to construct the C-alpha statistics for testing <5.7.1.1> against <5.7.1.5>, say,

$$CA_0 = u_0^{*\prime}P_{2n}'(\alpha^*)(\Sigma_0^{*-1} \otimes P_X)P_{2n}(\alpha^*)u_0^*, \qquad \text{<5.7.4.1>}$$

and for equations <5.7.1.2> against <5.7.1.5>,

$$CA_1 = u_0^{*\prime}P_{3n}'(\alpha^*)(\Sigma_0^{*-1} \otimes P_X)P_{3n}(\alpha^*)u_0^*, \qquad \text{<5.7.4.2>}$$

where α^* summarises the dependence of P_{2n} and P_{3n} on the parameter estimators of the model <5.7.1.1>; $P_{2n}(\alpha^*)$ is defined by equation <5.5.1.4>:

$$P_{2n}(\alpha^*) = I_{mn} - (I_m \otimes X\Pi_0^*)H[H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)H]^{-1} \times H'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')$$

and similarly, $P_{3n}(\alpha^*)$ is defined by

$$P_{3n}(\alpha^*) = I_{mn} - (I_m \otimes X\Pi_0^*)K[K'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)K]^{-1} \times K'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X').$$

These projection matrices satisfy

$$P_{2n}(I_m \otimes P_X) = (I_m \otimes P_X)P_{2n}, \quad P_{3n}(I_m \otimes P_X) = (I_m \otimes P_X)P_{3n},$$

so that the difference of the statistics <5.7.4.1> and <5.7.4.2> is

$$\begin{aligned} CA_0 - CA_1 &= u_0^{*\prime}(P_{2n}'(\alpha^*) - P_{3n}'(\alpha^*))(\Sigma_0^{*-1} \otimes P_X)u_0^* \\ &= u_0^{*\prime}(P_{2n}(\alpha^*) - P_{3n}(\alpha^*))'(\Sigma_0^{*-1} \otimes P_X)(P_{2n}(\alpha^*) - P_{3n}(\alpha^*))u_0^*, \end{aligned} \quad \text{<5.7.4.3>}$$

and in fact,

$$P_{2n}(\alpha^*) - P_{3n}(\alpha^*) = P_{2n}(\alpha^*)(I_m \otimes X\Pi_0^*)K[K'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)K]^{-1} \times K'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X');$$

pre- and post-multiplying this expression by $I_m \otimes P_X$ leaves it unchanged, so that <5.7.4.3> equals

$$u_0^{*\prime}P_{2n}'(\alpha^*)(\Sigma_0^{*-1} \otimes X\Pi_0^*)K[K'(\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X'X\Pi_0^*)K]^{-1}K' \times (\Sigma_0^{*-1} \otimes \Pi_0^{*\prime}X')P_{2n}(\alpha^*)u_0^*,$$

which is precisely the expression for CA given in equation <5.5.1.6>. So, given the use of initial estimators based on inefficient estimation of the null hypothesis model of equation <5.7.1.1>,

$$CA = CA_0 - CA_1.$$

But, there is no reason in general why such common initial estimators for models <5.7.1.1> and <5.7.1.2> should be used:

thus, one concludes that asymptotic equivalence of CA and CA_0-CA_1 is more likely in practice.

<u>5.7.5.</u> For Wald test statistics, one has to distinguish between those statistics based on a minimum chi-squared principle, discussed for the simultaneous equations model in subsection 5.6.2., and those based on the traditional Wald test principle, as in subsection 5.6.1., and shown to be related to the version of the minimum chi-squared based Wald test statistic based on the structural parameter estimators of the alternative hypothesis model. The reason is that differences of symmetric Wald statistics of the form <5.6.2.3> cannot be converted directly into asymmetric Wald statistics like <5.6.1.1> or <5.6.3.3>; one would only obtain an asymptotic equivalence.

Similarly, only when three-stage or constrained indirect least squares estimators are used will the difference in statistics like <5.6.1.1> and <5.6.3.2>, but appropriate for testing <5.7.1.1> against <5.7.1.5>, and <5.7.1.2> against <5.7.1.5>, equal the overall Wald statistic for tests of <5.7.1.1> against <5.7.1.2>. The reasons for this are the same as given in subsection 5.6.6. .

5.8. Limited Information Tests

5.8.1. In this Chapter, and in Chapter 3, the emphasis has been on "system" or "full information" estimation of a simultaneous equations model, and on "system-wide" inference: in many instances, however, it is natural to consider estimation and inference equation by equation, which normally requires the use of a limited information estimator. The test statistics in this section will generally be based on the LIML estimator presented in section 3.7.; the modifications necessary for the use of other estimators will be noted as appropriate.

It will be helpful to recall the type of simultaneous equations model within which LIML estimation was discussed in section 3.7.: there is one over-identified equation, the first, described by equation <3.7.2.2>,

$Z_1 c_{0.1} = u_{0.1}$,

but where the vector of structural parameters $c_{0.1}$ is restricted by equation <3.7.2.3>,

$c_{0.1} = H_{11}\gamma_{.1} + h_{.1}$;

$u_{0.1}$ has variance $\sigma_{0.11}$. The remaining $m - 1$ equations of the model are in reduced form, from equation <3.7.2.1>,

$Y_2 = X\Pi_{02} + V_{02}$;

the vector vec Π_{02} is treated as the vector of free structural parameters so that one can express this collection of equations in the form

$(I_{m-1} \otimes Z_1)g_{0.2} = u_{0.2}$,

$g_{0.2} = H_{22}\pi_{02} + h_{.2}$:

see equation <3.7.2.5>.

This description is taken to represent the model ruling under the the null hypothesis; under the alternative hypothesis, one would write for the first equation,

$Z_1c_{1.1} = u_{1.1}$,

with

$c_{1.1} = K_{11}\delta_{.1} + k_{.1}$.

Here, K_{11} is a known $(m+k_1) \times q_{11}$ matrix of full column rank, $\delta_{.1}$ is $q_{11} \times 1$, and $k_{.1}$ is a known $(m+k_1) \times 1$ vector.

One could therefore describe the null and alternative hypothesis models as

H_0': $c_{0.1} = H_{11}\gamma_{.1} + h_{.1}$ <5.8.1.1>

H_1': $c_{1.1} = K_{11}\delta_{.1} + k_{.1}$, <5.8.1.2>

or by supposing that the parameter vector $\delta_{.1}$ is obtained from $\gamma_{.1}$, so that the null hypothesis model is a more restricted version of the alternative hypothesis model. That is, suppose that under the null hypothesis,

$\delta_{.1} = L_{11}\gamma_{.1} + r_{.1}$,

the notation matching that used in the "system" case: see equation <3.1.3.1>. Here, L_{11} is a $q_{11} \times q_{01}$ matrix of full column rank, and $r_{.1}$ a known $q_{11} \times 1$ vector. Then, the null and alternative hypotheses may be described as

H_0: $\delta_{.1} = L_{11}\gamma_{.1} + r_{.1}$ <5.8.1.3>

H_1: $\delta_{.1} \neq L_{11}\gamma_{.1} + r_{.1}$. <5.8.1.4>

5.8.2. It will be convenient to discuss tests of this null hypothesis against the alternative when the first equation is just-identified under the alternative hypothesis, leaving the extension to the case where <5.8.1.2> is over-identified to follow.

The Likelihood Ratio statistic is easily obtained in this context: from equation <3.7.5.4>,

$$n^{-1}l_n(y;\tilde{\theta}) = -m(s + \tfrac{1}{2}n^{-1}) - \tfrac{1}{2}\log\det\hat{\Omega} + \tfrac{1}{2}\log(1 - \tilde{\nu}),$$

under the null hypothesis, whilst under the just-identified alternative hypothesis,

$$n^{-1}l_n(y;\hat{\theta}) = -m(s + \tfrac{1}{2}n^{-1}) - \tfrac{1}{2}\log\det\hat{\Omega},$$

so that

$$-2(l_n(y;\tilde{\theta}) - l_n(y;\hat{\theta})) = -n\log(1 - \tilde{\nu}) \overset{a}{\sim} \chi^2_{k_1 - q_{01}}. \quad \text{<5.8.2.1>}$$

This is equivalent to the statistic given by Anderson and Rubin [1949] and Koopmans and Hood [1953], apart from the slight change of notation.

When the alternative hypothesis model is over identified, let the smallest characteristic root of the determinental equation appropriate for the alternative hypothesis model denoted by $\tilde{\nu}_1$; then, the Likelihood Ratio statistic is essentially that given by Kadane [1974], equation <5.2.2.5> above, but with a notation change to

$$LR = -n\,[\log(1 - \tilde{\nu}) - \log(1 - \tilde{\nu}_1)]: \quad \text{<5.8.2.2>}$$

5.8.3. For a Lagrange Multiplier statistic suitable for a test of hypothesis <5.8.1.3> against hypothesis <5.8.1.4>

when the latter hypothesis is just-identified, there is a slight complication. In the "FIML" derivation of section 3.7. of the estimator $\tilde{\gamma}_{.1}$, there arose a system vector of Lagrange Multipliers, $\tilde{\tau}_2$, associated with the relationship between vec Π_0 and $\gamma_{.1}$, vec Π_{02}, whilst a subvector $\tilde{\tau}_{2.1}$ of $\tilde{\tau}_2$ arose from the Anderson and Rubin-based derivation od subsection 3.7.7.. However, the arguments of subsection 3.7.5. show that the system Lagrange multiplier $\tilde{\tau}_2$ depends directly on the single equation Lagrange multiplier $\tilde{\tau}_{2.1}$, so that a Lagrange Multiplier statistic based on $\tilde{\tau}_2$ will coincide with one based on $\tilde{\tau}_{2.1}$. Only the latter type of statistic will be presented.

From equation <3.7.4.3>,

$$\tilde{\tau}_{2.1} = -\tilde{\sigma}_{1.11}n^{-1}X'\tilde{u}_{0.1},$$

so that

$$LM = \tilde{\sigma}^{-1}_{0.11}\tilde{u}'_{0.1}XP_{1n}(\tilde{\alpha})(X'X)^{-1}P'_{1n}(\tilde{\alpha})X'\tilde{u}_{0.1},$$

where $P_{1n}(\tilde{\alpha})$ is defined by equation <3.7.6.5>; however, equation <3.7.6.4> shows that

$$P'_{1n}(\tilde{\alpha})X'\tilde{u}_{0.1} = X'\tilde{u}_{0.1},$$

so that

$$LM = \tilde{\sigma}^{-1}_{0.11}\tilde{u}'_{0.1}P_X\tilde{u}_{0.1} \qquad \text{<5.8.3.1>}$$

$$= n(\tilde{u}'_{0.1}\tilde{u}_{0.1})^{-1}(\tilde{u}'_{0.1}P_X\tilde{u}_{0.1})$$

$$= n(\tilde{c}'_{0.1}Z'_1Z_1\tilde{c}_{0.1})^{-1}(\tilde{c}'_{0.1}Z'_1P_XZ_1\tilde{c}_{0.1})$$

$$= n\tilde{\nu}, \qquad \text{<5.8.3.2>}$$

by equation <3.7.3.12>, since

$$\tilde{u}_{0.1} = Z_1\tilde{c}_{0.1}.$$

This is quite a striking result, and may be compared with the form of the Likelihood Ratio statistic of equation <5.8.2.1>,

$$LR = - n \log (1 - \tilde{v})$$

$$\simeq n\tilde{v}$$

using the expansion

$$- \log (1 - x) \simeq x, \quad 0 < x < 1.$$

There is also quite a strong analogy between this Lagrange Multiplier statistic and the two-stage least squares version of equation <5.2.3.1> given by Basmann [1960].

A direct form of this Lagrange Multiplier statistic follows from equation <5.8.3.1>, by regressing the residual vector $\tilde{u}_{0.1}$ on the matrix X: from this, the Lagrange Multiplier statistic can also be obtained as n times the uncentred R^2 of this regression. This particular form matches the full system Lagrange Multiplier statistic <5.4.2.2> very well.

If $\tilde{u}_{0.1}$ and $\tilde{\sigma}_{0.11}$ are replaced by the two-stage least squares residual vector $\hat{u}_{0.1}$ and variance estimator $\hat{\sigma}_{0.11}$ obtained by minimising

$$(K_{11}\delta_{.1} + k_{.1})'Z_1'P_XZ_1(K_{11}\delta_{.1} + k_{.1})$$

subject to

$$\delta_{.1} = L_{11}\gamma_{.1} + r_{.1},$$

when the alternative hypothesis model is just identified, one can show that equation <5.8.3.1> is of the correct form, with the numerator being the two-stage least squares residual squared norm,

$(H_{11}\hat{\gamma}_{.1} + h_{.1})'Z_1'P_XZ_1(H_{11}\hat{\gamma}_{.1} + h_{.1}),$

since in this case, the two-stage least squares estimators of $\gamma_{.1}$ and $\delta_{.1}$, $\hat{\gamma}_{.1}$ and $\tilde{\delta}_{0.1}$, satisfy

$$\hat{\delta}_{0.1} = L_{11}\tilde{\gamma}_{.1} + r_{.1},$$

$$K_{11}\hat{\delta}_{0.1} + k_{.1} = H_{11}\hat{\gamma}_{.1} + h_{.1}.$$

Because the alternative hypothesis is just-identified, $\hat{\gamma}_{.1}$ is also a single equation constrained indirect least squares estimator; the Lagrange Multiplier statistic is then equivalent to that given in equation <5.2.3.2>, allowing for the changed nature of the restrictions and the alternative hypothesis.

To obtain a Lagrange Multiplier statistic for the case where the alternative hypothesis model is overidentified, one would formally maximise

$$n^{-1}l_n(y;\phi)$$

as defined by equation <3.7.5.1>, but with Ω_0, Π_0 replaced by Ω_1, Π_1, subject to

$$c_{1.1} = K_{11}\delta_{.1} + k_{.1}$$

and

$$\delta_{.1} = L_{11}\gamma_{.1} + r_{.1}.$$

A repetition of the analysis of subsection 3.7.7., compared with the discussion of equation <5.4.1.4>, will enable one to argue, if only by analogy (to avoid tedious algebraic argument) that the appropriate Lagrange Multiplier statistic is

$$LM = \tilde{\sigma}_{0.11}^{-1}\tilde{u}_{0.1}'X\tilde{Q}_0K_{11}[K_{11}'\tilde{Q}_0'X'X\tilde{Q}_0K_{11}]^{-1}K_{11}'\tilde{Q}_0'X'\tilde{u}_{0.1}. \quad \text{<5.8.3.3>}$$

In this expression, the numerator is the explained sum of squares of the least squares regression of $\tilde{u}_{0.1}$ on $X\tilde{\Omega}_0 K_{11}$, whilst the denominator is n^{-1} times the uncentred "total sum of squares" of the regression, so that one could represent this statistic in the form nR^2.

The idea of using the difference of Lagrange Multiplier statistics for testing H_0' of equation <5.8.1.1> against a just identified alternative, and H_1' of equation <5.8.1.2> against the same just-identified alternative is interesting here, mainly because of the structure of the component Lagrange Multiplier statistics. The statistic of equation <5.8.3.2> may be expressed as

$LM_0 = n\tilde{v}$,

whilst the corresponding statistic for testing <5.8.1.2> against a just identified alternative will be

$LM_1 = n\tilde{v}_1$:

compare subsection 5.8.2. . Then, by the general results of section 5.7.,

$LM \overset{a}{\approx} LM_0 - LM_1 = n(\tilde{v} - \tilde{v}_1)$

which is directly comparable to the Likelihood Ratio statistic of equation <5.8.2.2>.

There is a direct two-stage least squares version of the statistic <5.8.3.3>, which simply replaces $\tilde{u}_{0.1}$ and $\tilde{\sigma}_{0.11}$ by the two-stage least squares values $\hat{u}_{0.1}$ and $\hat{\sigma}_{0.11}$, and $\tilde{\Omega}_0$ by the least squares estimator $\hat{\Omega}$; this Lagrange Multiplier statistic can be obtained from the explained squared norm of a two-stage least squares regression of $\hat{u}_{0.1}$ on Z_1K_{11}. The "difference" form of this statistic can be written in the form of equation <5.2.3.2> as

$$\hat{\sigma}^{-1}_{0.11}RSN(H_0)-\hat{\sigma}^{-1}_{1.11}RSN(H_1) = \hat{\sigma}^{-1}_{0.11}\hat{u}'_{0.1}P_X\hat{u}_{0.1}-\hat{\sigma}^{-1}_{1.11}\hat{u}'_{1.1}P_X\hat{u}_{1.1},$$

using two-stage least squares estimators of the null and alternative hypothesis models respectively; it will collapse to the statistic of equation <5.2.3.2> if the same variance estimator is used overall.

5.8.4. The C-alpha statistic for testing the null hypothesis of equation <5.8.1.3> against the alternative of <5.8.1.4>, when the latter corresponds to an overidentified equation, is related to the system C-alpha statistic of <5.5.1.6> in the same way that the single equation Lagrange Multiplier statistic is related to the system Lagrange Multiplier statistic of equation <5.4.1.4>.

The appropriate C-alpha statistic is

$$CA = n\sigma^{*-1}_{0.11}u^{*\prime}_{0.1}P'_{11n}(\alpha^*)X\Omega^*_0K_{11}[K'_{11}\Omega^{*\prime}_0X'X\Omega^*_0K_{11}]^{-1} \times K'_{11}\Omega^{*\prime}_0X'P_{11n}(\alpha^*)u^*_{0.1}, \qquad \text{<5.8.4.1>}$$

where $\sigma^*_{0.11}$ and $u^*_{0.1}$ stem from the inefficient initial estimators. The projection $P_{11n}(\alpha^*)$ is defined by

$$P_{11n}(\alpha^*) = I_n - X\Omega^*_0H_{11}[H'_{11}\Omega^{*\prime}_0X'X\Omega^*_0H_{11}]^{-1}H'_{11}\Omega^{*\prime}_0X'.$$

The statistic can be obtained by a regression of

$P_{11n}(\alpha^*)u^*_{0.1}$

on $X\Omega^*_0K_{11}$.

When the alternative hypothesis is just identified, $\Omega^*_0K_{11}$ is square and non-singular, producing

$$CA = n\sigma^{*-1}_{0.11}u^{*\prime}_{0.1}P'_{11n}(\alpha^*)P_XP_{11n}(\alpha^*)u^*_{0.1},$$

which is obtainable by a two-stage least squares regression of $u^*_{0.1}$ on $X\Omega^*_0H_{11}$, or by a least squares regression of $P_{11n}(\alpha^*)u^*_{0.1}$ on X.

A "difference" form for the C-alpha statistic <5.8.4.1> will follow from the arguments of subsection 5.7.4., with the same conditions for exact decomposition or asymptotic equivalence with the overall C-alpha statistic.

<u>5.8.5.</u> The natural Wald test principle would seem to be based on the argument of subsection 5.6.3., regressing the alternative hypothesis LIML estimator $\tilde{\delta}_{.1} - r_{.1}$ on L_{11} in the metric of

$$K'_{11}\tilde{\Omega}'_1X'X\tilde{\Omega}_1K_{11},$$

to produce the statistic

$$W = n\tilde{\sigma}^{-1}_{1.11}(\tilde{\delta}_{.1}-L_{11}\gamma^{\ddagger}_{.1}-r_{.1})'K'_{11}\tilde{\Omega}'_1X'X\tilde{\Omega}_1K_{11}(\tilde{\delta}_{.1}-L_{11}\gamma^{\ddagger}_{.1}-r_{.1}),$$

for the case where the alternative hypothesis model is over identified. If $\tilde{\delta}_{.1}$ is replaced by the two-stage least squares estimator $\hat{\delta}_{.1}$, and $\tilde{\Omega}_1$ by the least squares estimator $\hat{\Omega}$, then $\delta^{\ddagger}_{.1}$ is actually the two-stage least squares estimator $\hat{\gamma}_{.1}$, and the numerator of the statistic has the form

$$RSN(\mathbf{H}_0) - RSN(\mathbf{H}_1),$$

RSN being the two-stage least squares residual squared norm.

When the alternative hypothesis is just-identified, $\tilde{\Omega}_1 \equiv \hat{\Omega}$, and $\gamma^{\ddagger}_{.1}$ is actually the constrained indirect and two-stage least squares estimator of $\gamma_{.1}$; in addition, $RSN(\mathbf{H}_1) \equiv 0$, so that the statistic has the same form as the Lagrange Multiplier statistic of equation <5.8.3.1>, apart from the use of two-stage least squares residuals. Note that the alternative hypothesis variance estimator is used in this Wald statistic, whereas the null hypothesis estimator is used in the Lagrange Multiplier statistic: this is the usual situation encountered in Wald and Lagrange Multiplier statistics in the general linear model.

5.9. Conclusions

5.9.1. The observations made in section 2.13. concerning the properties of Likelihood Ratio, Lagrange Multiplier, Wald and C-alpha statistics in the general situation dicussed in that Chapter naturally remain valid the the specific case of the simultaneous equations model.

The relationship between the C-alpha and the minimum chi-squared Wald test noted in subsection 2.13.1., arising from the use of different approximations to the same quantity,

$\tilde{\theta} - \theta^*, \quad I_n^{-1}(\theta^*)D_\theta l_n(y;\theta^*)$

becomes a little different in the simultaneous equations model. When the alternative hypothesis model is just-identified, the test statistics are identical in form (see equation <5.5.2.3>), apart from the differing choices of the reduced form covariance matrix estimator: $\hat{\Omega}$ for the Wald, and Ω^* for the C-alpha. When the alternative hypothesis is over-identified, the C-alpha statistic of equation <5.5.1.6> uses the structural form residuals,

$u_0^* = (I_m \otimes Z_1)g_0^*$

associated with the inefficient initial estimator γ^*, whilst the Wald statistic of equation <5.6.3.3> uses the "fitted values" $X\tilde{Q}_1$ of Z_1 from estimation of the alternative hypothesis model, together with the linearised minimum chi-squared estimator of g_0:

$(I_m \otimes X\tilde{Q}_1)g_0^{\ddagger}$.

Since the relationship

$\beta = \lambda(\alpha)$

is linear in this case, no linearisation is required.

The characteristics of the various test principles of requiring estimation only under the null hypothesis, or only under the alternative hypothesis, or both (e.g. Lagrange Multiplier and C-alpha; minimum chi-squared Wald; Likelihood Ratio) are preserved in the simultaneous equations model case, as are the symmetry and asymmetry properties. These properties refer to the need for knowledge of the structural form of the alternative hypothesis model in order to conduct a test of the (additional) over-identifying restrictions imposed to obtain the null hypothesis model. One can argue that the specification-misspecification test properties are also preserved: all of the test statistics are misspecification tests when the alternative hypothesis is just-identified, and specification tests when it is over-identified.

The use of "difference statistics", that is, test statistics formed as the differences of test statistics for testing an over-identified null hypothesis against any just-identified alternative, adds another aspect to this question: such statistics are likely to appeal to the user who wishes to establish that neither the over-identified null hypothesis, nor the over-identified alternative hypothesis models he wishes to consider suffer from misspecification, relative to a just-identified model, in the sense that all of the overidentifying restrictions are false, before considering a test of his null hypothesis model against the over-identified alternative hypothesis. Such a user is really performing a multiple comparison, and would be wise to allow for this in his choice of test size. However, it is a pragmatic approach in model building, and raises again the issue of the users' objectives: is estimation undertaken solely for purposes of inference, or is inference an adjunct of estimation in "learning from one's data" ?

When inference is an adjunct of estimation, one is likely to prefer a statistic using only estimation under the null hypothesis, or only under the alternative hypothesis, if only one over-identified model is to be estimated, (such as the Lagrange Multiplier, C-alpha or Wald), whilst if both null and alternative hypothesis models are to be estimated, the Likelihood Ratio or a difference statistic seems appropriate. Even when the alternative hypothesis is just-identified, the Lagrange Multiplier statistics require a further regression for their calculation: the C-alpha and minimum chi-squared Wald statistics are produced free as a by-product of the estimation process. This remains true for over-identified alternatives.

5.9.2. In the introduction to this Chapter, a question arose as to what is being tested by "tests of over-identifying restrictions"; for, when the alternative hypothesis model, in general,

$\theta = \phi(\beta)$

is just-identified, the specific structural restrictions implied by $\phi(\beta)$ are irrelevant from the point of view of testing an overidentified null hypothesis defined by adding the restrictions

$\beta = \lambda(\alpha)$

to $\theta = \phi(\beta)$. That is, the restrictions $\beta = \lambda(\alpha)$ are whatever is necessary to go from

$\theta = \phi(\beta)$

to

$\theta = \theta(\alpha)$,

i.e.,

$\beta = \phi^{-1}\theta(\alpha)$,

since $\phi(.)$ is invertible in the just-identified case.

This question arose in a different way in section 5.2., in surveying the literature on tests of over-identifying restrictions: the conventional view, as expressed by Fisher and Kadane [1974], as well as other authors, is that <u>any</u> set of overidentifying restrictions is being tested. This is essentially the viewpoint expresed in the preceding paragraph. In practice, an investigator may wish to test a number of restrictions which are over-identifying, for substantive theoretical reasons: this has to be distinguished from the preceding case in that the null hypothesis model plus the specific over-identifying restrictions amount to a specific choice of alternative hypothesis model. For, if the alternative hypothesis model is supposed just-identified,

$\beta = \lambda(\alpha) \equiv \phi^{-1}\theta(\alpha)$

(for some ϕ) defines

$\theta = \phi(\beta)$.

It is as well to recall the basic supposition that the null hypothesis model is generated from the alternative hypothesis model by further restriction: so, when the alternative hypothesis is over-identified, these questions do not arise.

Again, one reaches the enforced conclusion that what is being tested depends on the investigator's objectives, and his desire for self-consistency, in using hypothesis testing to learn from the data: an investigator who uses a symmetric test pays the price of "generality" if his null hypothesis is rejected. Which restrictions are the source of the rejection ?

5.9.3. Another aspect of this debate concerns the way in which the various test statistics rely on different quantities: for example, the Likelihood Ratio statistic depends on the estimated reduced form covariance matrices, whilst the Lagrange Multiplier statistic depends on the reduced form or structural form residuals of the null hypothesis model. In fact, in the case of a just-identified alternative hypothesis model, the Lagrange Multiplier, minimum chi-squared Wald, and C-alpha statistics depend only on reduced form quantities, whilst for an over-identified alternative, they depend on structural form quantities.

The way in which reduced form quantities are used in the various test statistics (for a just identified alternative) suggests that "restrictions on the reduced form" implied by the over-identified null hypothesis model are being tested: this can be seen formally, anyway, since π_1 is unrestricted under the alternative hypothesis, and

$$\pi_1 = \theta_2(\gamma)$$

can be equivalently represented (at least, locally) as

$$f(\pi_1) = 0$$

for a suitable vector function $f(.)$.

When the alternative hypothesis is over-identified, the same point can be made, even though all the test statistics except the Likelihood Ratio statistic can be expressed solely in terms of structural form quantities: under the alternative hypothesis,

$$f_1(\pi_1) = 0,$$

and under the null hypothesis, additional restrictions must be satisfied,

$$f_2(\pi_1) = 0.$$

The analysis has been conducted in a "constraint parameter" framework, rather than this constraint equation framework (as used by Byron [1974] and Wegge [1978]) simply because the estimation and distributional aspects of FIML are much simpler. It is worth recalling from subsection 5.2.4. that despite the "symmetry" of the constraint equation framework, both Byron and Wegge were forced to "solve out" a just-identified set of parameters for the purpose of conveniently obtaining the limit distributions on which test statistics can be based.

<u>5.9.4.</u> One issue raised by Fisher and Kadane [1974] for a test of an over-identified model against a just-identified alternative which has not been discussed here is whether the alternative hypothesis can contain some unidentified models (see the end of subsection 5.2.2.). An answer will be attempted in the next Chapter, in connection with estimation and inference in unidentified models.

<u>5.9.5.</u> The discussion of inference for a model consisting of a single over-identified equation, and the remaining equations in reduced form, using LIML estimation was intended to provide a link with some of the traditional tests of over-identifying restrictions (Anderson and Rubin [1949], Koopmans and Hood [1953]), and also to provide a link with the more common single equation estimators like two-stage least squares, with its associated variants of the "asymptotic tests" discussed in subsection 5.2.3. . In addition, the approach adopted parallels naturally that obtained in the full system case.

More specifically, the Likelihood Ratio test statistics have exactly the form given by Anderson and Rubin [1949] and Kadane [1974] for rather specific types of restrictions, even though quite general linear restrictions have been imposed here. The same is true of the Lagrange Multiplier statistics discussed in subsection 5.8.3., whilst the Wald test principle of subsection 5.8.5. leads naturally to the use of two-stage or constrained indirect least squares estimators.